



Journal of
***Risk and Financial
Management***

Special Issue Reprint

Editorial Board Members' Collection Series

Journal of Risk and Financial Management

Edited by
Thanasis Stengos

mdpi.com/journal/jrfm



Editorial Board Members' Collection
Series: Journal of Risk and
Financial Management

Editorial Board Members' Collection Series: Journal of Risk and Financial Management

Guest Editor

Thanasis Stengos



Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Guest Editor

Thanasis Stengos

Department of Economics and Finance

University of Guelph

Guelph, ON

Canada

Editorial Office

MDPI AG

Grosspeteranlage 5

4052 Basel, Switzerland

This is a reprint of the Special Issue, published open access by the journal *Journal of Risk and Financial Management* (ISSN 1911-8074), freely accessible at: https://www.mdpi.com/journal/jrfm/special_issues/9MH46N7246.

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, Firstname, Firstname Lastname, and Firstname Lastname. Article Title. <i>Journal Name</i> Year , Volume Number, Page Range.

ISBN 978-3-7258-3413-6 (Hbk)

ISBN 978-3-7258-3414-3 (PDF)

<https://doi.org/10.3390/books978-3-7258-3414-3>

© 2025 by the authors. Articles in this book are Open Access and distributed under the Creative Commons Attribution (CC BY) license. The book as a whole is distributed by MDPI under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

Contents

About the Editor vii

Thanasis Stengos
Editorial Board Members' Collection Series: *Journal of Risk and Financial Management*
Reprinted from: *J. Risk Financial Manag.* **2025**, 18, 88, <https://doi.org/10.3390/jrfm18020088> . . . 1

Richard J. Cebula
Economic Freedom, Budget Deficits, and Perceived Risk from Larger National Debt-to-GDP Ratios: An Exploratory Analysis of Their Real Interest Rate Effects
Reprinted from: *J. Risk Financial Manag.* **2024**, 17, 469, <https://doi.org/10.3390/jrfm17100469> . . . 5

William A. Barnett, JongSoo Lee and Naowar Mohiuddin
Constructing Divisia Monetary Aggregates for the Asian Tigers
Reprinted from: *J. Risk Financial Manag.* **2024**, 17, 435, <https://doi.org/10.3390/jrfm17100435> . . . 16

Susana Álvarez-Otero and Emma Álvarez-Valle
The NGDOs Efficiency: A PROMETHEE Approach
Reprinted from: *J. Risk Financial Manag.* **2024**, 17, 382, <https://doi.org/10.3390/jrfm17090382> . . . 50

Anastasios G. Malliaris, Mary Malliaris and Mark S. Rzepczynski
One Man's Bubble Is Another Man's Rational Behavior: Comparing Alternative Macroeconomic Hypotheses for the US Housing Market
Reprinted from: *J. Risk Financial Manag.* **2024**, 17, 349, <https://doi.org/10.3390/jrfm17080349> . . . 76

H. D. Vinod
May 2024 Buy-Sell Guide for Dow Jones 30 Stocks and Modified Omega Criterion
Reprinted from: *J. Risk Financial Manag.* **2024**, 17, 343, <https://doi.org/10.3390/jrfm17080343> . . . 97

Markus Brueckner
Action-Based Fiscal Consolidations and Economic Growth
Reprinted from: *J. Risk Financial Manag.* **2024**, 17, 194, <https://doi.org/10.3390/jrfm17050194> . . . 110

Mark J. Kamstra and Ruoyao Shi
Testing and Ranking of Asset Pricing Models Using the GRS Statistic
Reprinted from: *J. Risk Financial Manag.* **2024**, 17, 168, <https://doi.org/10.3390/jrfm17040168> . . . 136

Kohei Marumo and Steven Li
Downside Risk in Australian and Japanese Stock Markets: Evidence Based on the Expectile Regression
Reprinted from: *J. Risk Financial Manag.* **2024**, 17, 189, <https://doi.org/10.3390/jrfm17050189> . . . 161

Sonal Sahu, José Hugo Ochoa Vázquez, Alejandro Fonseca Ramírez and Jong-Min Kim
Analyzing Portfolio Optimization in Cryptocurrency Markets: A Comparative Study of Short-Term Investment Strategies Using Hourly Data Approach
Reprinted from: *J. Risk Financial Manag.* **2024**, 17, 125, <https://doi.org/10.3390/jrfm17030125> . . . 178

Yuxin Zhang, Rajiv Garg, Linda L. Golden, Patrick L. Brockett and Ajit Sharma
Segmenting Bitcoin Transactions for Price Movement Prediction
Reprinted from: *J. Risk Financial Manag.* **2024**, 17, 128, <https://doi.org/10.3390/jrfm17030128> . . . 192

Masoumeh Fathi, Klaus Grobys and James W. Kolari
On the Realized Risk of Foreign Exchange Rates: A Fractal Perspective
Reprinted from: *J. Risk Financial Manag.* **2024**, 17, 79, <https://doi.org/10.3390/jrfm17020079> . . . 209

Nektarios Gavrilakis and Christos Floros

Volatility and Herding Bias on ESG Leaders' Portfolios Performance

Reprinted from: *J. Risk Financial Manag.* **2024**, 17, 77, <https://doi.org/10.3390/jrfm17020077> . . . **223**

Haim Levy and Moshe Levy

Option Pricing with the Logistic Return Distribution

Reprinted from: *J. Risk Financial Manag.* **2024**, 17, 67, <https://doi.org/10.3390/jrfm17020067> . . . **245**

Annette Hofmann and Peter Zweifel

Can Investment Views Explain Why People Insure Their Cell Phones But Not Their Homes?—A New Perspective on the Catastrophe Insurance Puzzle

Reprinted from: *J. Risk Financial Manag.* **2024**, 17, 30, <https://doi.org/10.3390/jrfm17010030> . . . **261**

Joseph K. W. Fung, F. Y. Eric Lam and Yiuman Tse

The Impact of ESG Rating on Hedging Downside Risks: Evidence from a Weight-Tilted Hang Seng Index

Reprinted from: *J. Risk Financial Manag.* **2024**, 17, 57, <https://doi.org/10.3390/jrfm17020057> . . . **273**

Miramir Bagirov and Cesario Mateus

A Survey of Literature on the Interlinkage between Petroleum Prices and Equity Markets

Reprinted from: *J. Risk Financial Manag.* **2024**, 17, 40, <https://doi.org/10.3390/jrfm17010040> . . . **290**

Thilini V. Mahanama, Abootaleb Shirvani, Svetlozar Rachev and Frank J. Fabozzi

The Financial Market of Indices of Socioeconomic Well-Being

Reprinted from: *J. Risk Financial Manag.* **2024**, 17, 35, <https://doi.org/10.3390/jrfm17010035> . . . **338**

John Fraser, Rob Quail and Betty Simkins

What's Wrong with Enterprise Risk Management?

Reprinted from: *J. Risk Financial Manag.* **2024**, 17, 274, <https://doi.org/10.3390/jrfm17070274> . . . **358**

Mark Joshi, Oh Kang Kwon and Stephen Satchell

Monte Carlo Sensitivities Using the Absolute Measure-Valued Derivative Method

Reprinted from: *J. Risk Financial Manag.* **2023**, 16, 509, <https://doi.org/10.3390/jrfm16120509> . . . **369**

Baixiao Liu, John J. McConnell and Andrew Schrowang

The Effect of Short-Sale Restrictions on Corporate Managers

Reprinted from: *J. Risk Financial Manag.* **2023**, 16, 486, <https://doi.org/10.3390/jrfm16110486> . . . **393**

Gilles Saint-Paul

Knowledge Sharing and Cumulative Innovation in Business Networks

Reprinted from: *J. Risk Financial Manag.* **2024**, 17, 137, <https://doi.org/10.3390/jrfm17040137> . . . **416**

About the Editor

Thanasis Stengos

Thanasis Stengos joined the Department of Economics at the University of Guelph in 1984, where he held a University Research Chair from 2004 to 2019. He received his B.Sc. and M.Sc. in Economics from the London School of Economics and Ph.D. from Queen's University. He currently serves as Associate Editor of Empirical Economics and he is the Editor-in-Chief of the *Journal of Risk and Financial Management* and co-editor of the *Review of Economic Analysis*. His research has been published in many journals including the *Review of Economic Studies*, *European Economic Review*, *International Economic Review*, *The Economic Journal*, *The Econometrics Journal*, *Journal of Monetary Economics*, *Journal of Econometrics*, *Econometric Theory*, *The Review of Economic and Statistics*, *Journal of Applied Econometrics*, *Journal of Business and Economic Statistics* and *Journal of Economic Growth*.



Editorial

Editorial Board Members' Collection Series: *Journal of Risk and Financial Management*

Thanasis Stengos

Department of Economics and Finance, University of Guelph, Guelph, ON N1G2W1, Canada;
tstengos@uoguelph.ca

Recently, the *Journal of Risk and Financial Management* completed a Special Issue that included papers based on invitations from the Editorial Board Members on a variety of areas in applied financial economics and risk management. It contains a very diverse collection of papers spanning many areas of interest that the academic community is actively pursuing to promote research at the frontiers of the disciplines in question. The aim was to provide an avenue for networking and communication between *JRFM* and scholars in the field of financial and economic risk and management, and the final collection attests to the success of the Issue in reaching this aim.

Below, I have provided an overview of the papers included in this Special Issue in the order in which they were published.

The paper by Richard J. Cebula, "Economic Freedom, Budget Deficits, and Perceived Risk from Larger National Debt-to-GDP Ratios: An Exploratory Analysis of Their Real Interest Rate Effects", revisits the issue of the impact of government budget deficits on interest rates and the hypothesis that larger deficits raise interest rates, providing strong evidence that the real interest rate yield on 10-year Treasuries in Canada and the real interest rate yield on 10-year U.S. Treasury notes are increasing functions of the central government budget deficits in both Canada and the U.S. and the real interest rate yields on 10-year Treasuries in Canada and 10-year U.S. Treasury notes are both decreasing functions of economic freedom in Canada and the U.S., respectively. The paper by William A. Barnett, JongSoo Lee and Naowar Mohiuddin, "Constructing Divisia Monetary Aggregates for the Asian Tigers", constructs Divisia monetary aggregates for the "Asian Tigers"—Hong Kong (1999–2024), South Korea (2009–2024), Singapore (1991–2021), and Taiwan (2005–2024)—and assesses whether Divisia monetary aggregates can explain nominal GDP better than simple-sum money. The key finding demonstrates that Divisia indices respond more sensitively to economic shocks and Divisia monetary aggregates can benefit monetary policy analyses in these territories. The paper by Susana Álvarez-Otero and Emma Álvarez-Valle, titled "The NGDOs Efficiency: A PROMETHEE Approach", tries to establish some objective indicators that would allow the efficiency of the operation of non-governmental development organizations (NGDOs) to be evaluated, since knowledge of the management of their previous projects could contribute towards the improvement of their future achievements. The results of this research reveal that 44.6% of the NGDOs (33 out of the 74 studied) operate inefficiently, compared to 29.7% of the sample considered in the study which were found to be efficient.

The paper by Anastasios G. Malliaris, Mary Malliaris and Mark S. Rzepczynski, "One Man's Bubble Is Another Man's Rational Behavior: Comparing Alternative Macroeconomic Hypotheses for the US Housing Market", considers how competing macroeconomic hypotheses have been developed to explain the U.S. housing market and its possible bubble

Received: 20 January 2025

Accepted: 3 February 2025

Published: 5 February 2025

Citation: Stengos, T. (2025). Editorial Board Members' Collection Series: *Journal of Risk and Financial Management*. *Journal of Risk and Financial Management*, 18(2), 88. <https://doi.org/10.3390/jrfm18020088>

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

behavior. Using a neural network model as an atheoretical non-linear approach to capture the relative importance of alternative macro variables, the authors show that these hypotheses generate different macro relevance and reach different conclusions on the presence of bubbles, creating bubble identification issues. The paper by H. D. Vinod, “May 2024 Buy-Sell Guide for Dow Jones 30 Stocks and Modified Omega Criterion”, looks at recent monthly data to help guide long-term investors to buy or sell the 30 Dow Jones Industrial Average (DJIA) Index components and constructs six stock-picking algorithms to provide an average of their results using the past 474 months (39+ years) of monthly closing prices, ending in May 2024. The paper reports twelve sets of ranks for both out-of- and in-sample versions of the six algorithms, and it was found that averaging the twelve sets yields the top and bottom k stocks.

The paper by John Fraser, Rob Quail and Betty Simkins “What’s Wrong with Enterprise Risk Management?” considers the failure of companies to properly implement enterprise risk management (ERM) and the reasons behind this. As such, the study provides a base against which further future research can be performed as ERM best practices continue to evolve. The study by Markus Brueckner, “Action-Based Fiscal Consolidations and Economic Growth”, examines the hypothesis that action-based fiscal consolidations have a negative effect on GDP growth, which their findings strongly reject. The reason given is that the least squares estimates presented in previous studies suffer from negative reverse causality bias: GDP growth has a significant positive effect on both the likelihood and the magnitude of action-based fiscal consolidations, and to properly uncover the causal effects of action-based fiscal consolidations, one needs to rely on an instrumental variables approach. In their study, “Downside Risk in Australian and Japanese Stock Markets: Evidence Based on the Expectile Regression”, Kohei Marumo and Steven Li consider an expectile-based Value at Risk (EVaR) approach to evaluate the stock market indices of Australia and Japan. The findings suggest that both the Australian and Japanese stock markets are affected by their past development and international stock markets. Additionally, the ASX 200 index has significant impact on Nikkei 225 in terms of downside tail risk, while the impact of Nikkei 225 on ASX is not significant. In their paper, “Testing and Ranking of Asset Pricing Models Using the GRS Statistic”, Mark J. Kamstra and Ruoyao Shi assess the GRS statistic by providing the correct formula for it and the first proof of its F-distribution in the general multiple-factor case. Casual generalization of the Sharpe-ratio-based interpretation of the single-factor GRS statistic to the multiple-portfolio case makes asset pricing studies susceptible to using an incorrect formula. The authors illustrate the consequences of using the incorrect formulas as the ambiguity in the GRS leads to over-rejecting and mis-ranking asset pricing models. Gilles Saint-Paul in “Knowledge Sharing and Cumulative Innovation in Business Networks” shows that, if innovations are cumulative, making an invention publicly available to a network of firms may be valuable if the firm expects to benefit from future improvements made by other firms. A cooperative equilibrium where all innovations are made public is shown to exist under certain conditions and it is crucial not to deviate from this equilibrium regardless of another firm’s actions following a deviation. The greater the number of firms in the network, the more likely it is for a cooperative equilibrium to arise. When R&D efforts are endogenous, cooperative equilibria are associated with strategic complementarities between firms’ research efforts, which may lead to multiple equilibria. In “Segmenting Bitcoin Transactions for Price Movement Prediction”, Yuxin Zhang, Rajiv Garg, Linda L. Golden, Patrick L. Brockett and Ajit Sharma make use of standard machine learning models and blocks of publicly available transaction data to predict the direction of Bitcoin price movement. The paper shows that segmenting publicly available transactions based on investor typology helps achieve a higher prediction accuracy compared to the existing Bitcoin price movement prediction models used in the

literature. This transaction segmentation highlights the role of investor types in impacting financial markets. The paper by Sonal Sahu, José Hugo Ochoa Vázquez, Alejandro Fonseca Ramírez and Jong-Min Kim, “Analyzing Portfolio Optimization in Cryptocurrency Markets: A Comparative Study of Short-Term Investment Strategies Using Hourly Data Approach”, investigates portfolio optimization methodologies and short-term investment strategies in the context of the cryptocurrency market, focusing on ten major cryptocurrencies from June 2020 to March 2024. The empirical analysis reveals significant volatility, skewness and kurtosis in cryptocurrencies, highlighting the need for sophisticated portfolio management techniques, and the main finding is that the Kurtosis Minimization methodology used in the paper consistently outperforms other optimization strategies, especially in shorter-term investment horizons, delivering optimal returns to investors. The paper emphasizes the importance of dynamic portfolio management, stressing the necessity of regular rebalancing in the volatile cryptocurrency market. Masoumeh Fathi, Klaus Grobys and James W. Kolari in “On the Realized Risk of Foreign Exchange Rates: A Fractal Perspective” follow Mandelbrot by taking a fractal perspective indicating that power laws offer an alternative to the lognormal in terms of goodness-of-fit tests. The main finding is that the foreign exchange rate market is far riskier than previously believed and, by implication, research documented in the enormous body of literature that draws conclusions from variance analyses stands on shaky grounds.

In “Volatility and Herding Bias on ESG Leaders’ Portfolios Performance”, Nektarios Gavrilakis and Christos Floros analyze the factor loadings given by the CAPM, the Fama–French three (FF3) and the five-factor model (FF5), and test the performance and the validity of adding two more factors (volatility and dispersion of returns) to the FF5 factor model of European index-based ESG leaders’ portfolios. They find that European ESG leaders’ portfolios are biased towards a large cap and value stocks with robust operating profitability and against aggressive investments and that European and Global ESG leaders’ portfolios tilt towards volatility and herding bias. The paper by Haim Levy and Moshe Levy, “Option Pricing with the Logistic Return Distribution”, derives analytic option pricing formulas for an underlying asset with a logistic return distribution. These formulas are simple and elegant and employ exactly the same parameters as B&S. The logistic option pricing formula fits empirical option prices much better than B&S, providing an explanatory power comparable to much more complex models with a larger number of parameters. In “The Impact of ESG Rating on Hedging Downside Risks: Evidence from a Weight-Tilted Hang Seng Index”, Joseph K. W. Fung, F. Y. Eric Lam and Yiuman Tse examine the return performance and resilience to market volatility of the recently introduced environment, social/sustainable and governance (ESG) weight-tilted Hang Seng index compared to its parent, the Hang Seng index. The results show that stocks with high ESG ratings are less susceptible to trading pressures triggered by volatility-induced turnovers providing a significant incremental information on the emerging market for ESG-related equity products in Hong Kong. In “A Survey of Literature on the Interlinkage between Petroleum Prices and Equity Markets”, Miramir Bagirov and Cesario Mateus provide an extensive review of a plethora of empirical studies in this strand of the literature. By scrutinizing over 190 papers published from 1983 to 2023, the survey reveals various research themes and points to diverse findings that are sector- and country-specific and contingent on the methodologies, data frequencies and time horizons employed. Petroleum price changes and shocks exert direct or indirect effects dictated by the level of petroleum dependency across sectors and the country’s position as a net petroleum exporter or importer, and the interlinkages between these factors tend to display a time-varying nature and sensitivity to major market events. The paper by Thilini V. Mahanama, Abootaleb Shirvani, Svetlozar Rachev and Frank J. Fabozzi, “The Financial Market of Indices of Socioeconomic Well-Being”, discusses how

financial economic theory and its quantitative tools can be applied to create socioeconomic indices and develop a financial market for the so-called “socioeconomic well-being indices”. While environmental, social and governmental (ESG) financial markets address socioeconomic issues, the focus in the paper is broader, encompassing national citizens’ well-being. The findings propose a financial market model for well-being indices, encouraging the financial industry to adopt and trade these indices as mechanisms to manage and protect them against downturn risks. In “Can Investment Views Explain Why People Insure Their Cell Phones But Not Their Homes?—A New Perspective on the Catastrophe Insurance Puzzle”, Annette Hofmann and Peter Zweifel examine the consistently missing demand for catastrophe insurance and for coverage of other low-probability–high-consequence risks is often referred to as the catastrophe insurance puzzle. This paper points out that the underlying rationale may be that individuals interpret insurance contracts with a low payout probability as an investment with a negative expected net present value. The main suggestion is that combining a low-probability with a high-probability loss insurance contract may be a way to incentivize individuals to purchase catastrophe risk coverage. In “Monte Carlo Sensitivities Using the Absolute Measure-Valued Derivative Method”, Mark Joshi, Oh Kang Kwon and Stephen Satchell propose an alternative approach to the recently adopted computationally intensive measure-valued differentiation (MVD) method, called the absolute measure-valued differentiation (AMVD) method, which expresses the derivative of the transition density at each simulation step as a single density rather than a linear combination. It is computationally more efficient than the MVD method and can result in sensitivity estimates with lower variance and analytic and numerical examples are provided to compare the variance in the sensitivity estimates of the AMVD method against alternative methods. Finally, the paper by Baixiao Liu, John J. McConnell and Andrew Schrowang, “The Effect of Short-Sale Restrictions on Corporate Managers”, studies the effect of short selling on corporate managers from 2002 to 2010 and finds that, when deciding whether to abandon value-reducing acquisition attempts during the program, managers of pilot firms, whose stocks are less subject to short selling impediments, are more sensitive to stock price changes than managers of nonpilot firms. These results indicate that, despite their dislike of short sellers, managers believe that the level of informativeness from capital markets is superior when short sellers are less impeded.

The above papers exemplify the diverse scope of the areas of research covered in this Special Issue and offer a strategy for further high-caliber collections on this topic that would be on the cutting edge of current research.

Conflicts of Interest: The author declares no conflict of interest.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

Economic Freedom, Budget Deficits, and Perceived Risk from Larger National Debt-to-GDP Ratios: An Exploratory Analysis of Their Real Interest Rate Effects

Richard J. Cebula

Economics Department, University of Tennessee at Knoxville, Knoxville, TN 37996, USA; rcebula@utk.edu

Abstract: Since the early 1980s, there have been a number of principally empirical studies of the impact of government budget deficits on interest rates that have typically tested the hypothesis that larger deficits raise interest rates. However, in more recent years, this topic has received far less attention. Accordingly, this study seeks to “update” the findings of such studies and to do so for the dominant North American economies of Canada and the U.S. Furthermore, in the pursuit of further insights into interest rates, the present study also investigates an effectively heretofore overlooked variable that arguably also might influence interest rates, namely, economic freedom. Finally, given the increased upward trend of government debt (relative to GDP) in recent years in Canada and the U.S., this study investigates the interest rate impact of rising national debt-to-GDP ratios. For the 1995–2024 period (and also in one estimate for the 1985–2001 period), this exploratory study finds compelling evidence (1) that the real interest rate yield on 10-year Treasuries in Canada and the real interest rate yield on 10-year U.S. Treasury notes are *increasing* functions of the central government budget deficits in both Canada and the U.S., respectively, and (2) the real interest rate yields on 10-year Treasuries in Canada and 10-year U.S. Treasury notes are both *decreasing* functions of *economic freedom* in Canada and the U.S., respectively. On the other hand, regarding the impact of a higher national debt-to-GDP ratio on the real ten-year Treasury yield, there is only very mixed support for an impact, with support for its impact coming from the Canadian estimates but no support whatsoever coming from the U.S. estimates.

Keywords: economic freedom; government budget deficit; elevated national debt-to-GDP ratios; real interest rate yield

Citation: Cebula, Richard J. 2024. Economic Freedom, Budget Deficits, and Perceived Risk from Larger National Debt-to-GDP Ratios: An Exploratory Analysis of Their Real Interest Rate Effects. *Journal of Risk and Financial Management* 17: 469. <https://doi.org/10.3390/jrfm17100469>

Academic Editor: Thanasis Stengos

Received: 21 September 2024

Revised: 15 October 2024

Accepted: 16 October 2024

Published: 17 October 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Federal government budget deficits have been studied both theoretically, e.g., within the context of a government budget constraint, and empirically (using a variety of econometrics tools), over the years with the deficit expressed/measured in a variety of forms. These budget deficit forms have included those of the total budget deficit per se, the structural budget deficit (Darrat 2000), the cyclical budget deficit (Barth et al. 1984, 1985), and even the total budget deficit net of debt service payments (Zahid 1988; Cebula and Rhodd 1993). Studies of the economic impacts and ramifications of such government budget deficits on interest rates (both real and nominal), as well as on the term structure of interest rates, inflation, real GDP growth, unemployment rates, and employment growth, along with other economic or financial yardsticks, have been undertaken for a number of nations. These nations include the U.K., Italy, France, Germany, the U.S., Canada, Greece, and Spain, and reach across much of the globe to a variety of developing nations. Regardless of the form of the budget deficit being researched, however, a very common finding is that higher budget deficits, typically expressed as a percentage of GDP, tend to elevate intermediate and longer-term interest rates (Al-Saji 1993; Barth and Bradley 1989; Cebula and Koch 1994; Cebula et al. 2016; Choi and Holmes 2014; Cotton 2021; Cross 2022; Cukierman and

Meltzer 1989; Zahid 1988; Darrat 2000; Ewing and Yanochik 1999; Findlay 1990; Gale and Orszag 2003, 2004; Gisse 1999; Hoelscher 1986; Holloway 1986; Johnson 1992; McMillan 1986; Swamy et al. 1990; Tanzi 1985; Quayes and Jamal 2007; Laubach 2009; Cotton 2021); however, these deficits are typically found to not affect short-term interest rate yields.¹ Not surprisingly, there is reason to believe that rising interest rates exacerbate the cost of servicing the national debt (Cebula 2014; Cukierman and Meltzer 1989; Laubach 2009; Madura 2008), as well as complicating corporate governance in the realm of investment in new plant and equipment decisions, arguably making continued/renewed research into the deficit–interest rate issue more important.

This exploratory study has three primary objectives. Firstly, given that the impact of central government budget deficits on interest rates has received only limited attention in the scholarly literature during recent years, the first of the primary objectives of the present study is to empirically investigate and update, for the cases of the neighboring North American nations of Canada and the U.S. (whose economies are very highly interdependent), the impact of higher central government budget deficits (expressed as a percent of GDP) on (1) the ex post real interest rate yield on 10-year Canadian Treasuries for the case of Canada and (2) the ex post real interest rate yield on 10-year U.S. Treasury notes for the case of the U.S. Secondly, another primary objective of this study is to help provide potentially useful, salient new information, unlike what is found in the extant related literature, regarding whether higher levels of *economic freedom* (The Heritage Foundation 2024) reduce the ex post real interest rate yield on 10-year Treasury obligations in Canada and in the U.S. Given that higher levels of economic freedom tend to yield greater financial market efficiency in a variety of ways (to be explained further in Section 2 below), it is hypothesized that the higher the level of economic freedom, the lower the real interest rate yield. Thirdly, the remaining primary objective of this study is to investigate the hypothesis that a higher national debt, relative to GDP (i.e., relative to the size of the economy) elevates the perceived risk of default on the principal and/or interest and therefore elevates the interest rate required in the financial marketplace to compensate for that perceived elevated risk. The exploratory time series analysis presented here uses annual data within a loanable funds framework to investigate the validity of these three hypotheses. The study initially considers the time period from 1995 through to 2024 and subsequently considers the period 1985–2001 (the latter made possible using economic freedom data derived by Stansel et al. (2023)).

2. The Framework and Study Objectives

To help identify factors influencing 10-year central government Treasury yields in Canada and the U.S., respectively, we focus on the ex post real interest rate yields on these financial instruments within the context of a loanable funds model (paralleling but extending somewhat those found in Barth 1991; Barth et al. 1985; Brandl 2020; Madura 2008). Since this study deals with two separate nations, the model is interpreted such that all of the variables considered below are associated with either Canada on the one hand or the U.S. on the other hand. Accordingly, in each of the loanable funds models, the ex post real yield on 10-year Treasury issues is, assuming all other bond markets are in equilibrium, determined by the following:²

$$D + CY + MY = S + DEFY \quad (1)$$

where D is private domestic demand for 10-year Treasury issues (in one of two forms: either (a) 10-year Canadian Treasuries or (b) 10-year U.S. Treasury notes). Next, S is the market supply for each nation of such 10-year Treasury issues. To reflect the open economy dimension of the model, we follow Klein (1978), Barth (1991), Barth et al. (1984), Barth et al. (1991), Cebula (2014), and others, who include a variable CY , which is the ratio for each nation of net financial capital inflows to GDP, whereas MY is the ratio of the M2 money stock to the level of GDP. $DEFY$ is the total central government (Canadian or U.S.) budget deficit, expressed as a percent of GDP in Canada or the U.S. The value of the ex post

real yield on these Canadian or U.S. Treasury notes ($EPR10$) is simply the nominal yield on 10-year Treasuries in each of these nations ($NOM10YLD$) minus that nation's actual inflation rate (P):

$$EPR10 = NOM10YLD - P \quad (2a)$$

Similarly, the ex post real yield on three-month Treasury bills is the nominal yield on those bills ($NOM3MTB$) minus the inflation rate:

$$EPR3 = NOM3MTB - P \quad (2b)$$

The total market demand for 10-year Treasury issues (debt) may be expressed as follows (Al-Saji 1993; Barth et al. 1984, 1985, 1991; Hoelscher 1986; Cebula 2014):

$$D = D(EPR10, EPR3, EFI, UNRATE, DEBTY) \quad (3)$$

such that

$$D_{EPR10} > 0, D_{EPR3} < 0, D_{EFI} > 0, D_{UNRATE} < 0, D_{DEBTY} < 0$$

Meanwhile, the supply of those 10-year Treasuries is expressed as follows:

$$S = S(EPR10, EPR3) \quad (4)$$

such that

$$S_{EPR10} < 0, S_{EPR3} > 0$$

where the variable $EPR10$ is the *ex post* real interest rate yield on 10-year Canadian Treasuries or U.S. Treasuries and $EPR3$ is the *ex post* real interest rate yield on 3-month Canadian or U.S. Treasury bills.

According to the model expressed above, the private sector demand for 10-year Treasury obligations is an increasing function of $EPR10$, *ceteris paribus*; this is because rational bond buyers prefer a higher real rate of return on their investment over a lower rate of return. On the other hand, Treasury note suppliers/issuers would supply/issue/offer to the financial market *fewer* high-grade municipal bonds in response to a higher $EPR10$ since such a condition would raise the debt service costs, *ceteris paribus*, and hence the total cost of their public investment and other outlays. However, if the cost of supplying more 3-month T-bills rises, they may issue more 10-year Treasuries instead, *ceteris paribus*. Next, the higher the ex post real interest rate yield on 3-month Treasury bills ($EPR3$), the lower the private sector demand for 10-year Treasury issues, as investors substitute such higher-paying 3-month issues for the longer-term and arguably riskier 10-year bonds,³ *ceteris paribus*.

It is hypothesized here that the private sector demand for Treasury notes is an increasing function of *economic freedom*, whose value ranges from a potential low of 0.0 to a potential high of 100.0.⁴ A higher level of economic freedom implies a diminished degree of government regulation, fewer governmental/agency rules, and generally fewer impediments to financial transactions involving bonds, equities, and ETFs in the marketplace. Additionally, greater economic freedom implies less interference in and lower pecuniary transactions costs in the financial decision process and in the subsequent execution process, as well as fewer impediments in the exercise of free judgment and free will of economic agents, be they individuals, households, non-profit corporations, or for-profit organizations, or other types of enterprise or institution. Arguably, then, greater economic freedom, measured here initially by the Heritage Foundation Economic Freedom Index, EFI , promotes more efficient economic and financial transactions in an economy. This presumably would be reflected in lower interest rate yields. Accordingly, it is hypothesized here that the *ex post* real interest rate yield on 10-year Canadian government Treasuries and the ex post real interest rate yield on 10-year U.S. Treasury notes are arguably decreasing functions of EFI .

Thus, the private sector demand for 10-year Treasury obligations is hypothesized to be positively related to the level of economic freedom, *ceteris paribus*. Hence, the first of

the three central hypotheses being investigated in this study of Canada and the U.S. is the following:

H1: *Ceteris paribus, the ex post real interest rate yield on 10-year Treasuries in Canada and the U.S. are decreasing functions of the economic freedom levels in Canada and the U.S., respectively.*⁵

The sequencing runs as follows: as *economic freedom* rises, private sector demand for Treasuries (*D*) rises, the Treasury price rises, and the Treasury yield declines.

In addition to the economic freedom index, there is the federal budget deficit, which is, as noted above, the second central focus of this study. Following the conventional wisdom (Carlson and Spencer 1975), it is hypothesized that the greater the central government budget deficit relative to its *GDP* (*DEFY*), the greater the upward pressure on interest rates generally (including *EPR10*) as a reflection of the central government's competing with the remainder of the economy for funds, *ceteris paribus* (Al-Saji 1993; Barth et al. 1984, 1985; Cebula and Koch 1994; Cebula 2014; Choi and Holmes 2014; Cukierman and Meltzer 1989; Engen and Hubbard 2004; Ewing and Yanochik 1999; Finlay et al. 2020; Gale and Orszag 2003, 2004; Gisse 1999; Gomme 2022; Hoelscher 1986; Johnson 1992; Klein 1974; Kolluri and Giannaros 1987; Swamy et al. 1990; Tanzi 1985; Quayes and Jamal 2007; Laubach 2009; Reinhart and Rogoff 2010; Cotton 2021). Hence, the second of the central hypotheses being investigated in this study of Canada and the U.S. is the following:

H2: *Ceteris paribus, the ex post real interest rate yield on 10-year Treasuries in Canada and the ex post real interest rate yield on 10-year U.S. Treasury notes are increasing functions of the central government budget deficit in Canada and in the U.S., respectively.*

Next, it is hypothesized that higher values for the control variable *DEBTY*, the ratio of the national debt to *GDP*, have been found in certain prior studies (Carlson and Spencer 1975; Klein 1978; Cukierman and Meltzer 1989; Cebula and Rhodd 1993; Engen and Hubbard 2004; Reinhart and Rogoff 2010) to have exerted upward pressure on interest rates in the U.S. Arguably this impact is a reflection of greater Treasury pressure on the financial markets as well as a reflection of the greater perceived riskiness of federal government debt management. The debt-to-*GDP* ratio, usually expressed in percentage terms, can be interpreted as an indicator of a nation's economic circumstances. Arguably, the higher a nation's *DEBTY*, the less likely that nation is to pay off that debt or the interest payable thereupon; alternatively stated, with a higher *DEBTY*, that nation's prospects of default may be regarded as increased. If the *DEBTY* rises, it may be that debt increases relative to the ability to discharge that debt. Alternatively stated, this is treated as hypothesis **H3**:

H3: *Ceteris paribus, the higher the DEBTY (the national debt-to-GDP ratio), the lower the demand for 10-year bonds due to perceived higher risk, ceteris paribus, and hence the lower their market prices, so that the interest rate yield on those bonds rises.*

To summarize, aside from the hypothesized relationships compounded above in **H1**, **H2**, and **H3**, the model initially consists of four *control* variables: *CY*, *MY*, *UNRATE*, and *EPR3*. The first of these is a net financial capital inflows variable, *CY*. Following the conventional wisdom, it is hypothesized that the greater the extent of net financial capital inflows relative to *GDP*, the greater the extent to which domestic debt issues are absorbed and hence the less the upward pressure on interest rate yields, *ceteris paribus*. Strong empirical support for this perspective is found for the U.S. in Cebula and Koch (1994). The variable *MY* is the ratio of *M2* money supply to the level of *GDP*, expressed as a percentage. The greater this ratio, the greater the degree to which the FED demands (absorbs), among other things, 10-year Treasury debt issues. Hence, it follows that *EPR10* is a decreasing function of *MY*, *ceteris paribus*. Logically, the higher the unemployment rate (*UNRATE*), the lower the demand for bonds, and hence the lower the price of bonds, and, consequently, ultimately the higher the yield on those bonds. Thus, it is expected here that the higher the

unemployment rate, the higher the ex post real interest rate yield. Similarly, the higher the real interest rate yield on three-month Treasury bills, the higher the demand for those bills and hence (due to substitution) the lower the demand for 10-year Treasuries, and hence the lower the price of the latter and the higher their yield.

Accordingly, the basic model is provided, as follows:

$$EPR10 = f(DEFY, EFI, UNRATE, EPR3, CY, MY, DEBTY) \tag{5}$$

Clearly, based upon the above discussion/modeling, it is hypothesized that

$$f_{DEFY} > 0, f_{EFI} < 0, f_{UNRATE} > 0, f_{EPR3} > 0, f_{CY} < 0, f_{MY} < 0, f_{DEBTY} > 0 \tag{6}$$

Furthermore, within the construct reflected in (5) and (6), the *principal* explanatory variables of interest are *DEFY*, *EFI*, and *DEBTY*, so as to reflect **H1**, **H2**, and **H3**, whereas the control variables initially are *UNRATE*, *EPR3*, *CY*, and *MY*.

3. Model and Data: The Case of Canada

Based upon the model expressed in Equations (5) and in (6) above, the following empirical model is to be considered initially:

$$EPR10_t = a_0 + a_1 DEFY_t + a_2 EPR3_t + a_3 UNRATE_{t-1} + a_4 CY_{t-1} + a_5 MY_{t-1} + a_6 EFI_{t-1} + a_7 DEBTY_{t-1} + a_8 AR(1) + a_9 AR(2) + \varepsilon_t \tag{7}$$

where *AR*(1) and *AR*(2) are the autoregressive terms (which are briefly discussed below) and ε_t is the stochastic error term. For Canada (and the U.S., which is considered in Section 3 of this study), the budget deficit, net capital flows variable, and M2 money supply are all scaled by GDP because the sizes of the budget deficit, net capital inflows, and money supply should be judged relative to the size of the economy (Cukierman and Meltzer 1989; Ostrosky 1990; Barth et al. 1984, 1985; Cebula and Koch 1994). As observed above, this study uses annual data from 1995 through to 2024. The data sources for Canada for the above model were the Bank of Canada (2023a, 2023b, 2023c), Statistics Canada (2023a, 2023b), and The Heritage Foundation (2024). For the interested reader, the VIFs (Variance Inflation Factors) are summarized in Table 1, where all are shown to be less than 10.0; therefore, the evidence implies that multicollinearity is not a serious statistical problem in this specification for Canada.⁶

Table 1. Variance inflation factors for initial estimates, Canada and the U.S.

Explanatory Variable	VIFs [Canada]	VIFs [U.S.]
DEFY	6.04	7.74
EPR3	7.01	6.51
CY	5.12	4.49
UNRATE	4.88	3.91
MY	5.49	5.92
EFI	3.17	3.28
DEBTY	6.11	7.75

Five of the seven explanatory variables in this model are lagged by one period. This lagging is intended to reflect the conventional wisdom that intermediate and long-term bond markets respond with a time lag to (a) unemployment rate changes; (b) monetary policy actions by the central bank, The Bank of Canada for Canada or the Federal Reserve for the U.S.; (c) economic freedom index changes; (d) changes in net financial capital inflows; and (e) growth in the debt-to-GDP ratio. Such lags have been analyzed in a variety of scholarly and academic outlets (Cukierman and Meltzer 1989; Feldstein and Eckstein 1970; Carlson and Spencer 1975; Klein 1974, 1978; Barth and Bradley 1989; Ostrosky 1990; Darrat 2000; Madura 2008; Reinhart and Rogoff 2010; Brandl 2020). However, the remaining two

explanatory variables, *DEFY* and *EPR3*, are both unlagged, which is consistent with much of the extant literature on deficits and intermediate and long-term interest rate yields (Al-Saji 1993; Barth et al. 1984, 1985; Cebula and Koch 1994; Choi and Holmes 2014; Cukierman and Meltzer 1989; Darrat 2000; Ewing and Yanochik 1999; Feldstein and Eckstein 1970; Gisseey 1999; Hoelscher 1986; Johnson 1992; Ostrosky 1990; Swamy et al. 1990; and Tanzi 1985).

Given that the dependent variable, the ex post real interest rate yield on 10-year Canadian Treasuries (*EPR10_t*), is contemporaneous with the variables *DEFY_t* and *EPR3_t*, the possibility of simultaneity bias arises. This problem can be addressed through an estimation by two-stage least squares (TSLS) after integrating/adopting appropriate instrumental variables into the analysis for each of the unlagged right-hand side/explanatory variables. Each of the instrumental variables in the estimate provided in Table 2 is lagged by two years. The instrumental variables adopted for the variable *DEFY_t* are the two-year lags of the average ex post real interest rate paid on long-term borrowing by businesses in Canada (*EPRBUS_{t-2}*) and the annualized percentage change in real GDP (*Y_{t-2}*); the instrument adopted for the variable *EPR3_t* is the two-year lag of the ex post real interest rate on 3-year Canadian Treasury bills (*EPR3Y_{t-2}*). The choice of instruments in this study is based (1) on the finding that for Canada the variables *EPRBUS_{t-2}* and *Y_{t-2}* are both highly correlated with the *DEFY_t* variable for Canada and the finding that the variable *EPR3Y_{t-2}* is highly correlated with *EPR3_t*, whereas (2) these two-period lagged instruments are uncorrelated with the error term in the system.

Table 2. AR/TSLS model estimate using a two-period lag on the instruments [Canada].

Dependent Variable: <i>EPR10</i>		
Explanatory Variables	Initial Estimate Coefficient	Second Estimate
Constant	22.15	23.66
<i>DEFY_t</i>	0.0455 ***	0.0461 ***
<i>EPR3_t</i>	0.6128 ***	0.6023 ***
<i>CY_{t-1}</i>	−0.0311	−0.0307
<i>UNRATE_{t-1}</i>	0.0982 *	0.0977 #
<i>MY_{t-1}</i>	−0.0428 ***	−0.0426 ***
<i>EFI_{t-1}</i>	−0.2646 ***	−0.2646 *
<i>DEBTY_{t-1}</i>	0.049 **	0.0425 **
<i>EPR30_{t-1}</i>		0.1423 ***
AR(1)	0.1212 ***	0.1226 ***
AR(2)	−0.5003 ***	0.4878 ***
Initial Estimate:		
J-Statistic = 24.32 **, DW = 2.01		
Second Estimate:		
J-Statistic = 24.21 **, DW = 2.05		

Newey–West heteroscedasticity correction adopted: *** Statistically significant at the 1% level; ** statistically significant at the 2.5% level; * statistically significant at the 5% level; # statistically significant at the 10% level.

The estimates provided in this study involve autoregressive, i.e., AR(1) and AR(2), processes. The AR(1) and AR(2) models are of interest as a process for many time series applications. They are perhaps best applicable to time series that exhibit more volatile behavior, such as stock market indices, equity stock prices, and, as investigated in this study, interest rates (Lütkepohl 1993; Enders 1995). The autoregressive/TSLS, i.e., AR/TSLS, estimate of Equation (9) with the Newey and West (1987) heteroscedasticity correction adopted is reported below in Table 2, where all seven estimated coefficients exhibit the hypothesized signs, with five of the seven statistically significant at the 1% level and one statistically significant at the 5% level. Thus, it is revealed that the ex post real interest rate yield on 10-year Canadian Treasuries (*EPR10_t*) is an increasing function of *DEFY_t*, *UNRATE_{t-1}*, *DEBTY*, and *EPR3_t*, while being a decreasing function of *EFI_{t-1}* and *MY_{t-1}*, although not significantly a function of *CY_{t-1}*. The J-statistic is statistically significant at the

5% level, attesting to the exogeneity of the instrumental variables adopted (Sargan 1958). In any event, these empirical results are supportive of hypotheses **H1**, **H2**, and **H3**.

The empirical analysis summarized in Table 2 can be extended/expanded in a variety of ways. One such extension would be to add the ex post real interest rate yield on 30-year Canadian Treasury bonds, *EPR30*, to the list of explanatory variables, arguably by including it in the private sector demand for 10-year Canadian Treasuries, where its competition with 10-year Treasuries can be reflected. Clearly, ceteris paribus, an increase in *EPR30* would at the margin increase the demand for 30-year Canadian Treasuries and in the process act through a process of substitution to reduce the demand for 10-year Canadian Treasuries, such that the price on the latter would decline, thereby raising their yield. Hence, it is hypothesized that, ceteris paribus, the variable *EPR10* would be an increasing function of the variable *EPR30*.⁷ Accordingly, the following specification is now to be estimated:

$$EPR10_t = a_0 + a_1 DEFY_t + a_2 EPR3_t + a_3 UNRATE_{t-1} + a_4 CY_{t-1} + a_5 MY_{t-1} + a_6 EFI_{t-1} + a_7 DEBTY_{t-1} + a_8 EPR30_{t-1} + a_9 AR(1) + a_{10} AR(2) + \varepsilon_t \quad (8)$$

The AR/TSLS, estimate of Equation (8) is provided in Table 2, where it is shown that all eight of the coefficients on the explanatory variables exhibit the hypothesized signs, with four (one of which is *EPR30_{t-1}*) being statistically significant at the 1% level, two being statistically significant at the 5% level, and one being significant at somewhat beyond the 10% level. Only the coefficient on the variable *CY_{t-1}* fails (once again) to be statistically significant at the 10% level. These results are effectively compatible with those in Table 2. Of greatest relevance in this modest de facto robustness test, the coefficients on both *EFI_{t-1}*, *DEFY_t* and *DEBTY_{t-1}* are statistically significant, with the expected signs at the 5%, 1%, and 2% levels, respectively. These three results are strongly supportive of the **H1**, **H2**, and **H3** hypotheses, respectively, for the case of Canada.

4. Model and Data: The Case of the U.S.

Based initially upon the analytical modeling expressed in Equation (8) above, the following empirical model is to be estimated initially for the case of the U.S.:

$$EPR10_t = a_0 + a_1 DEFY_t + a_2 EPR3_t + a_3 \log UNRATE_{t-1} + a_4 CY_{t-1} + a_5 dMY_{t-1} + a_6 EFI_{t-1} + a_7 DEBTY_{t-1} + a_8 EPR30_{t-1} + a_9 AR(1) + a_{10} AR(2) + \varepsilon_t \quad (9)$$

This modified AR/TSLS specification for the U.S. parallels that shown in Equation (9) for Canada. The data for this specification for the U.S. were obtained from the following: FRED (2023); the Council of Economic Advisors (2014, 2018, 2023); and The Heritage Foundation (2024). As was the case for the model addressing the 10-year real interest rate yield in Canada, the Variance Inflation Factors (VIFs) for the U.S. for the variables in Equation (9) were estimated. These VIFs are all shown in Table 1, and each was found to be less than 10.0; consequently, a multicollinearity problem does not exist in the U.S. real interest rate estimation.⁸

The AR/TSLS estimate of the model for the U.S., is provided in Table 3. Seven of the eight estimated coefficients in Table 3 exhibit the hypothesized signs, with three being statistically significant at the 1% level, one being statistically significant at the 5% level, and one being significant at the 10% level. Once again, as in the Canadian case, the capital inflows variable is not statistically significant at even the 10% level. Unlike the results for Canada reported in Table 2, the estimated coefficients for the unemployment rate and debt-to-GDP variables in the U.S. case also fail to be statistically significant at the 10% level. The *J*-statistic is statistically significant at the 5% level, attesting to the exogeneity of the instrumental variables.

The estimation results found in Table 3 imply that the ex post real interest rate yield on 10-year U.S. Treasury notes is an increasing function of the federal government budget deficit (at the 1% level) while being a decreasing function of the economic freedom index (at beyond the 5% level). Hence, it follows that hypotheses **H1** and **H2** are once again strongly

supported by the empirical evidence, in this case within the context of the U.S. Indeed, the empirical results for the U.S. and Canada closely parallel each other, with the ex post real interest rate yield on 10-year U.S. Treasury notes ($EPR10_t$) also being an increasing function of $DEFY_t$, $EPR30_{t-1}$, and $EPR3_t$, while being a decreasing function of EFI_{t-1} and MY_{t-1} but not of CY_{t-1} . Thus, although hypotheses H1 and H2 continue to exhibit a strong affirmation, hypothesis (H3) finds no empirical support.⁹

Table 3. AR/TSLS model estimate using a two-period lag on the instruments [U.S.].

Dependent Variable: <i>EPR10</i>		
Explanatory Variables	Initial Estimate Coefficient	Alternative Estimate
Constant	13.31	−1.44
<i>DEFY_t</i>	0.1056 ***	0.0744 ***
<i>EPR3_t</i>	0.2581 ***	0.2362 ***
<i>CY_{t−1}</i>	−0.0843	0.0872
<i>Log UNRATE_{t−1}</i>	0.2412	0.3258
<i>dMY_{t−1}</i>	−0.1160 #	−0.0368
<i>EFI_{t−1}</i>	−0.1793 **	
<i>dEFNAI_{t−1}</i>		−0.0005 ***
<i>DEBTY_{t−1}</i>	−0.0136	−0.8766
<i>EPR30_{t−1}</i>	0.7203 ***	0.8087 ***
AR(1)	0.1128	0.4013 **
AR(2)	−0.1954	−0.0301
Initial Estimate:		
<i>J</i> -Statistic = 24.00 **, DW = 2.13		
Alternative Estimate:		
<i>J</i> -Statistic = 31.60 **, DW = 2.01		

Newey–West heteroscedasticity correction adopted: *** Statistically significant at the 1% level; ** statistically significant at the 2.5% level; * statistically significant at the 5% level; # statistically significant at the 10% level.

Consequently, to undertake this modest de facto robustness test of the model summarized in Table 3, the following specification is to be estimated for the U.S.:¹⁰

$$EPR10_t = a_0 + a_1 DEFY_t + a_2 EPR3_t + a_3 logUNRATE_{t-1} + a_4 CY_{t-1} + a_5 dMY_{t-1} + a_6 dEFNAI_{t-1} + a_7 dDEBTY_{t-1} + a_8 EPR30_{t-1} + a_9 AR(1) + a_{10} AR(2) + \varepsilon_t \tag{10}$$

It should be observed that this robustness test estimation adopts a different measure of economic freedom, namely, the index provided in *Economic Freedom of North America* by Stansel et al. (2023), $EFNAI_{t-1}$, which provides a measure of overall economic freedom that largely parallels computations for 1995–2004 by The Heritage Foundation (2024), EFI , but which provides data for the period 1985 through 2021.

The AR/TSLS estimate of Equation (10) is provided in Table 3, along with the estimate of Equation (9). As revealed in this table, six of the eight estimated coefficients on the explanatory variables exhibit the hypothesized signs, with four (including the control variable $EPR30_{t-1}$, as well as $DEFY_t$, $EFNAI_{t-1}$, and $EPR3_t$) being statistically significant at the 1% level. As for the variables CY_{t-1} , MY_{t-1} , $DEBT_{t-1}$, and $UNRATE_{t-1}$, their respective t-statistics are not significant at even the 10% level. Of greater relevance to this study, the coefficients on $DEFY_t$ and $EFNAI_{t-1}$ are highly statistically significant with the expected signs and therefore supportive of hypotheses H1 and H2, respectively, for the case of the U.S. However, within this context of the economic freedom measure derived by Stansel et al. (2023), the coefficient on the debt-to-GDP variable, $DEBT_{t-1}$, fails to be statistically significant. Hence, there is further evidence that in the U.S. hypothesis H3 is empirically unsupported. Arguably, this lack of statistical significance on behalf of H3 reflects, however subtly, the role of the U.S. dollar as the reserve currency of most of the world and the fact that the U.S. is the world’s largest economy.

5. Conclusions

This empirical undertaking has sought to provide updated and new insights, albeit perhaps exploratory, to our understanding of the factors that influence the *ex post* real interest rate yield on government-issued debt, namely, 10-year Treasuries. To provide updated information, the study period adopts annual data for the period 1995–2024 (and then, in *one* estimate, 1985–2021). Two neighboring and highly integrated nations, Canada and the U.S., are studied. Three factors are focused upon: (1) most prominently, the heretofore effectively ignored financial market impact of *economic freedom* [as computed by The Heritage Foundation (2024) on the one hand and the Fraser Institute (Stansel et al. 2023) on the other hand], (2) the central/federal government budget deficit, expressed as a percentage of GDP, and (3) the oft-neglected central government debt-to-GDP ratio, expressed as a percentage.

Accordingly, three central hypotheses are investigated, with the AR/TSLS results provided in Table 2 for Canada and in Table 3 for the U.S. As shown in these tables, higher levels of economic freedom act to *reduce* the real interest rate yield on 10-year Treasury notes in both Canada and the U.S. Arguably, this is the most interesting contribution of the study. However, in both Canada and the U.S., higher central government budget deficits relative to GDP are found to raise the real interest rate yield on 10-year Treasuries. On the other hand, there is only mixed evidence regarding the interest rate impact of a higher debt-to-GDP ratio. Thus, the estimation results provided here consistently support only two of the three central hypotheses in this study for Canada and the U.S.

Future scholarly research involving the subject matter in the present study could take the form of investigating alternative *ex post* real interest rates, such as those for 30-year Treasuries, 20-year Treasuries, and 5-year or 7-year Treasuries. Short-term Treasury debt instruments and/or the term structure of interest rates could also be investigated. Moreover, a longer study period could be investigated. Perhaps longer lags on the instruments could be considered as well. More elaborate modeling could also be undertaken, possibly including a variety of financial market factors. Furthermore, more elegant econometrics tools could be applied, such as cointegration. Finally, perhaps an alternative to the national debt-to-GDP variable can be identified as a better reflection of risk, e.g., the bank failure rate.

Funding: This research received no external funding.

Data Availability Statement: Data are all accessible from identified publicly available sources. Email Requests for data from author can be sent via email.

Conflicts of Interest: The author declares no conflict of interest.

Notes

- ¹ However, there are a few exceptions. For example, Barth et al. (1984, 1985) have found the budget deficit in the U.S. to positively impact 3-month Treasury bills.
- ² Loanable funds models have been used extensively in the scholarly literature, e.g., Al-Saji (1993), Barth et al. (1984, 1985, 1991), Hoelscher (1986), and Cebula (2014).
- ³ This risk is referred to as “interest rate risk” (Cecchetti 2006).
- ⁴ The most recent overall EFI for Canada is 72.4 while that for the U.S. is 70.1.
- ⁵ Research underlying this study reveals no prior effort to explore economic freedom’s acting to reduce real interest rates, whether *ex post* real or *ex ante* real.
- ⁶ For the interested reader, the P-P and ADF tests both confirm that all of the variables in (7) and (8) are stationary in levels.
- ⁷ See Hoelscher (1986) for an example of the substitution between different U.S. Treasury issues, namely, 3 month T-bills and 10-year T-notes.
- ⁸ For the interested reader, the P-P and ADF tests both confirm that the only variable in this specification that is non-stationary in levels is MY_{t-1} ; all of the remaining variables in Equation (9) are stationary in levels.
- ⁹ Note that, paralleling our treatment of investigating hypotheses **H1**, **H2**, and **H3** for Canada, we introduce the *ex post* real interest rate yield on 30-year U.S. Treasuries (*EPR30*) into the private sector demand for 10-year U.S. Treasuries. This specification clearly implies that, for the case of the U.S., a rise in the value of *EPR30* should result in a higher demand for 30-year U.S. Treasuries but results at the same time in at least some degree of substitution of those 30-year issues for 10-year issues. Thus, it is *hypothesized* that through competition between these two Treasury debt issues, an elevated *EPR30* leads to an elevated *EPR10*, *ceteris paribus*.

- ¹⁰ For the interested reader, the P-P and SADF tests both confirm that the only variables in this specification that are non-stationary in levels are MY_{t-1} and $EFNA_{t-1}$; all of the remaining variables in Equation (9) are stationary in levels.

References

- Al-Saji, Amer K. 1993. Government budget deficits, nominal and ex ante real long-term interest rates in the UK, 1960: 1–1990: 2. *Atlantic Economic Journal* 21: 71–77. [CrossRef]
- Bank of Canada. 2023a. Selected Benchmark Bond Yields. Retrieved from Bank of Canada. Available online: <https://www.bankofcanada.ca/rates/interest-rates/canadian-bonds/#benchmark> (accessed on 2 March 2024).
- Bank of Canada. 2023b. Selected Treasury Bill Yields. Retrieved from Bank of Canada. Available online: <https://www.bankofcanada.ca/rates/interest-rates/t-bill-yields/> (accessed on 2 March 2024).
- Bank of Canada. 2023c. Summary of Key Monetary Policy Variables. Retrieved from Bank of Canada. Available online: <https://www.bankofcanada.ca/rates/indicators/key-variables/key-inflation-indicators-and-the-target-range/> (accessed on 2 March 2024).
- Barth, James R. 1991. *The Great Savings and Loan Debacle*. Washington, DC: American Enterprise Institute.
- Barth, James R., and Michael Bradley. 1989. Evidence on real interest rate effects of money, deficits, and government spending. *Quarterly Review of Economics and Business* 29: 49–57.
- Barth, James R., George R. Iden, and Frank S. Russek. 1984. Do federal deficits really matter? *Contemporary Economic Policy* 3: 79–95. [CrossRef]
- Barth, James R., George R. Iden, and Frank S. Russek. 1985. Federal borrowing and short-term interest rates: Comment. *Southern Economic Journal* 52: 554–59. [CrossRef]
- Barth, James R., George R. Iden, Frank S. Russek, and Mark Wohar. 1991. *The Effect of Federal Budget Deficits on Interest Rates and the Composition of Domestic Output*. Washington, DC: Urban Institute.
- Brandl, Michael W. 2020. *Money, Banking, Financial Markets & Institutions*. Boston: Cengage Learning.
- Carlson, Keith M., and Roger W. Spencer. 1975. Crowding out and its critics. *Federal Reserve Bank of St. Louis Review* 12: 1–15. [CrossRef]
- Cebula, Richard J. 2014. An empirical inquiry into the impact of federal budget deficits and the ex-ante real interest rate yield on long-term treasury notes. *Journal of Public Finance and Public Choice* 32: 83–97. [CrossRef]
- Cebula, Richard J., and James V. Koch. 1994. Federal budget deficits, interest rates, and international capital flows: A further note. *The Quarterly Review of Economics and Finance* 34: 117–20. [CrossRef]
- Cebula, Richard J., and Rupert G. Rhodd. 1993. A note on budget deficits, debt service payments, and interest rates. *Quarterly Review of Economics and Finance* 33: 439–35. [CrossRef]
- Cebula, Richard J., Wendy Gillis, S. Cathy McCrary, and Don Capener. 2016. New evidence on the impact of economic conditions and banking legislation on the bank failure rate in the US, 1970 to 2014. *Journal of Financial Economic Policy* 8: 364–76. [CrossRef]
- Cecchetti, Stephen G. 2006. *Money, Banking, and Financial Markets*. New York: McGraw Hill-Irwin.
- Choi, Daniel F. S., and Mark J. Holmes. 2014. Budget deficits and real interest rates: A regime-switching reflection on Ricardian equivalence. *Journal of Economics and Finance* 38: 71–83. [CrossRef]
- Cotton, Christopher D. 2021. *Debt, Deficits, and Interest Rates*. Research Department. Boston: Federal Reserve Bank of Boston, pp. 1–31.
- Council of Economic Advisors. 2014. *Economic Report of the President*; Washington, DC: U.S. Government Printing Office.
- Council of Economic Advisors. 2018. *Economic Report of the President*; Washington, DC: U.S. Government Printing Office.
- Council of Economic Advisors. 2023. *Economic Report of the President*; Washington, DC: U.S. Government Printing Office.
- Cross, Phillip. 2022. *Do Budget Deficits Matter? Essays on the Implications of Deficits and Debt*. Vancouver: Fraser Institute, chp. 2.
- Cukierman, Alex, and Allan H. Meltzer. 1989. A political theory of government debt and deficits in a Neo-Ricardian framework. *American Economic Review* 79: 711–32.
- Darrat, Ali F. 2000. Do structural budget deficits impact interest rates? *Public Finance Review* 28: 194–97. [CrossRef]
- Enders, Walter. 1995. *Applied Econometric Time Series*. New York: Wiley.
- Engen, Eric M., and R. Glenn Hubbard. 2004. Federal government debt and interest rates. *NBER Macroeconomics Annual* 19: 83–138. [CrossRef]
- Ewing, Bradley T., and Mark Yanochik. 1999. Budget deficits and the term structure of nominal interest rates in Italy. *Applied Economics Letters* 6: 199–201. [CrossRef]
- Feldstein, Martin, and Otto Eckstein. 1970. The Fundamental Determinants of the Interest Rate. *The Review of Economics and Statistics* 52: 363–75. [CrossRef]
- Findlay, David W. 1990. Budget deficits, expected inflation, and short-term interest rates. *International Economic Journal* 50: 41–53.
- Finlay, Richard, Claudia Seibold, and Michelle Xiang. 2020. *Government Bond Market Functioning and COVID-19*. Sydney: Reserve Bank of Australia, pp. 11–20.
- FRED. 2023. Available online: <https://fred.stlouisfed.org> (accessed on 31 January 2024).
- Gale, William G., and Peter R. Orszag. 2003. Economic effects of sustained budget deficits. *National Tax Journal* 56: 463–85. [CrossRef]
- Gale, William G., and Peter R. Orszag. 2004. Budget deficits, national savings, and interest rates. *Brookings Papers on Economic Activity* 34: 101–210. [CrossRef]
- Gissey, William. 1999. Net treasury borrowing and interest-rate changes. *Journal of Economics and Finance* 23: 23–29. [CrossRef]
- Gomme, Paul. 2022. US fiscal policy during and after the Coronavirus. *Canadian Journal of Economics* 55: 358–78. [CrossRef]
- Hoelscher, Gregory. 1986. New evidence on deficits and interest rates. *Journal of Money, Credit and Banking* 18: 1–17. [CrossRef]

- Holloway, Thomas M. 1986. The cyclically adjusted federal and federal debt: Revised and updated estimates. *Survey of Current Business* 66: 11–17.
- Johnson, Colleen F. 1992. An empirical note on interest-rate equations. *Quarterly Review of Economics and Finance* 32: 141–47.
- Klein, John J. 1974. The comparative statics of monetary policy targets. *Economic Notes* 6: 115–30.
- Klein, John J. 1978. *Money and the Economy*, 4th ed. New York: Harcourt, Brace, and Jovanovich.
- Kolluri, Bharat R., and Demetrios S. Giannaros. 1987. Budget deficits and short-term real interest rate forecasting. *Journal of Macroeconomics* 9: 109–25. [CrossRef]
- Laubach, Thomas. 2009. New evidence on the interest rate effects of budget deficits and debt. *Journal of the European Economic Association* 7: 858–85. [CrossRef]
- Lütkepohl, Helmut. 1993. *Introduction to Multiple Time Series Analysis*. Berlin/Heidelberg: Springer.
- Madura, Jeff. 2008. *Financial Markets and Institutions*, 8th ed. Mason: Thomson Higher Learning.
- McMillan, W. Douglas. 1986. Federal deficits and short term interest rates. *Journal of Macroeconomics* 8: 405–22. [CrossRef]
- Newey, Whitney K., and Kenneth D. West. 1987. A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55: 703–08. [CrossRef]
- Ostrosky, Anthony L. 1990. Federal government budget deficits and interest rates. *Southern Economic Journal* 56: 802–804. [CrossRef]
- Quayes, Shakil, and A. M. M. Jamal. 2007. Budget deficits and interest rates: The US evidence since 1946. *The Singapore Economic Review* 52: 191–200. [CrossRef]
- Reinhart, Carmen M., and Kenneth S. Rogoff. 2010. Growth in a time of debt. *American Economic Review* 100: 573–78. [CrossRef]
- Sargan, John D. 1958. The estimation of economic relationships using instrumental variables. *Econometrica* 26: 393–415. [CrossRef]
- Stansel, Dean, José Torra, Fred McMahon, and Ángel Carrión-Tavárez. 2023. *Economic Freedom of North America, 2023*. Vancouver: Fraser Institute.
- Statistics Canada. 2023a. Table 26-10-0016-01 Balance of International Payments, Current Account and Capital Account (x1,000,000). Retrieved from Statistics Canada. Available online: <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610001601> (accessed on 25 February 2024).
- Statistics Canada. 2023b. Chart 1 Canadian General Government Deficit as a Percentage of Gross Domestic Product. Retrieved from Statistics Canada. Available online: <https://www150.statcan.gc.ca/n1/daily-quotidien/230323/cg-c001-eng.htm> (accessed on 25 February 2024).
- Swamy, Paravastu A. V. B., Bharat R. Kolluri, and Rao N. Singamsetti. 1990. What do regressions of interest rates on deficits imply? *Southern Economic Journal* 56: 1010–1028. [CrossRef]
- Tanzi, Vito. 1985. Fiscal deficits and interest rates in the United States: An empirical analysis, 1960–84. *IMF Staff Papers* 32: 551–76. [CrossRef]
- The Heritage Foundation. 2024. 2024 Index of Economic Freedom. Available online: <https://www.heritage.org/press/2021-index-economic-freedom-global-economic-freedom-remains-all-time-high-us-drops-all-time> (accessed on 25 February 2024).
- Zahid, Khan H. 1988. Government budget deficits and interest rates: The evidence since 1971, using alternative deficit measures. *Southern Economic Journal* 54: 725–31. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

Constructing Divisia Monetary Aggregates for the Asian Tigers

William A. Barnett ^{1,2}, JongSoo Lee ^{1,*} and Naowar Mohiuddin ¹

¹ Department of Economics, University of Kansas, Lawrence, KS 66045, USA; naowarm@ku.edu (N.M.)

² Center for Financial Stability, New York, NY 10036, USA

* Correspondence: jongsoo@ku.edu

Abstract: This study constructs Divisia monetary aggregates for the “Asian Tigers”—Hong Kong (1999–2024), South Korea (2009–2024), Singapore (1991–2021), and Taiwan (2005–2024)—and assesses whether Divisia monetary aggregates explain nominal GDP better than simple-sum money. Our findings demonstrate that Divisia indices respond more sensitively to economic shocks. For Hong Kong and Taiwan, narrow Divisia money provides the best explanations for fluctuations in nominal GDP. Our results suggest that Divisia monetary aggregates can be beneficial for monetary policy analysis in these territories and underscore the importance of further research into the empirical performance of Divisia monetary aggregates in macroeconomic prediction.

Keywords: divisia index; divisia monetary aggregates; vector error-correction model

1. Introduction

The “Barnett critique” highlights the superiority of Divisia monetary aggregation over the traditional simple-sum method of monetary aggregation (Barnett et al. 1992; Chrystal and MacDonald 1994; Belongia 1996). The simple-sum approach, which assumes that all financial components are perfect substitutes, is not consistent with economic theory. Even much earlier studies (Chetty 1969; Moroney and Wilbratte 1976; Boughton 1981) found that financial assets are less than perfect substitutes and advocated for different weights to be assigned to each component within monetary aggregates. While the simple-sum method has historically been used to measure spikes in the growth rate of the money supply, Barnett (2016) has demonstrated that these simple-sum spikes can be misleading. For example, after the end of the “monetarist experiment period”, Milton Friedman observed a seemingly alarming surge in simple-sum M2, which Barnett attributed to regulatory changes having no significant effect on the demand or supply for monetary services in the economy. Barnett’s research highlights that the Divisia index, based on well-established aggregation and index number theory, offers a more accurate measure of the money supply by avoiding the distortions and unusual spikes that are inherent in the simple-sum method.

The survey by Kumah (1989) indicated that the measurement of money in over 150 countries is limited to M1, M2, and M3, depending on the level of development or monetization of the financial system. A comparison of M2 and Lf (which is similar to M3 in other countries), and inflation rates in South Korea (hereafter Korea) from January 1998 to September 2022 reveals that these traditional simple-sum monetary aggregates could not adequately reflect the inflation trend or nominal GDP during the same period. This observation underscores Barnett’s (1980) criticisms of adding financial assets’ components without appropriately weighing each asset’s contribution to the monetary service flow. Barnett argued that the simple-sum monetary aggregate, which assumes equal weights for all components, cannot accurately measure the flow of monetary services. He advocated the well-known Divisia quantity index, which weights component growth rates with the value shares for each component, to provide a measure consistent with economic aggregation and index number theory.

Citation: Barnett, William A., JongSoo Lee, and Naowar Mohiuddin. 2024. Constructing Divisia Monetary Aggregates for the Asian Tigers. *Journal of Risk and Financial Management* 17: 435. <https://doi.org/10.3390/jrfm17100435>

Academic Editor: Thanasis Stengos

Received: 26 August 2024

Revised: 23 September 2024

Accepted: 24 September 2024

Published: 29 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

The superiority of the Divisia monetary aggregates has been increasingly recognized across many countries. Studies have shown that the Divisia aggregates provide better recession signals, more accurate money-demand predictions, and improved macroeconomic forecasts for the Euro Area and the United States (Hendrickson 2014; Serletis and Gogas 2014; Barnett and Gaekwad 2021; Brill et al. 2021). In Asia, Divisia aggregates have been found to offer more precise monetary signals during financial crises and better forecasts of economic activity, particularly in Singapore, India, and China, although their efficacy during the COVID-19 pandemic posed challenges (Barnett and Nguyen 2021; Barnett et al. 2022b; Sengupta et al. 2024). Further confirmation of the Divisia monetary aggregates' utility in enhancing monetary policy comes from studies in the UK and Saudi Arabia (Alkhareif and Al-Rasasi 2021; Belongia and Ireland 2021; Barnett et al. 2022a). Since 2010, research in territories like Germany, Indonesia, Israel, Pakistan, Taiwan, and Turkey supports the advantages of the Divisia aggregates (Sarwar et al. 2010; Puah and Hiew 2011; Chen and Nautz 2015; Benchimol 2016; Binner and Kelly 2017; Polat 2018). However, relevant studies about Korea, Taiwan, Hong Kong, and Singapore have become less common in recent times (see Appendix A Table A1).

In light of this fact, active research on Divisia monetary aggregates in various Asian countries is now underway (Jha and Longjam 2008; Barnett and Nguyen 2021). However, there has been a noticeable gap in research for Hong Kong, Korea, and Taiwan—territories that continue to use simple-sum monetary indicators. The only previously published research on Divisia monetary aggregates for Taiwan was Shih (2000), who constructed Divisia monetary aggregates for Taiwan and evaluated the empirical performance of those aggregates against the standards chosen by the central bank for the intermediate target variable in the conduct of monetary policy. Additionally, such research on Hong Kong, Korea, and Taiwan could enable a meaningful focus on applying Divisia monetary aggregates to all developing and developed Asian territories. Habibullah (1999) states that the Divisia monetary aggregates are more effective in developed countries with market-oriented financial systems. Korea's recent reclassification as a developed country by the United Nations Conference on Trade and Development (UNCTAD) in July 2021 highlights the need to examine the usefulness and feasibility of Divisia monetary aggregates in this new context.

The term "Asian Tigers", also known as "Asian Dragons", collectively refers to Hong Kong, Korea, Singapore, and Taiwan, which experienced rapid economic growth from the 1960s to the 1990s (Figure 1). This term metaphorically compares these territories' dynamic and aggressive economic growth to the characteristics of tigers. There are two main reasons why our study focuses on the Asian Tigers. First, despite their high level of economic development, these territories still use simple-sum monetary aggregates and can benefit from exploring the Divisia approach. Second, the complexity and advanced nature of these territories' financial systems make them highly relevant for observing the effects of Divisia monetary measurement. For these reasons, research on Divisia monetary aggregates for the Asian Tigers is likely to provide significant academic and policy implications.

Our study aims to calculate and compare Divisia monetary aggregates with simple-sum aggregates for the Asian Tigers—Hong Kong, Korea, Singapore, and Taiwan. To achieve this, we utilize relevant data for each country's money supply and conduct a comparative analysis between the simple-sum and Divisia monetary aggregates. The study addresses two key research questions. First, can countries achieve more accurate analytical research outcomes using Divisia economic monetary aggregates instead of simple-sum accounting aggregates? Is Habibullah's (1999) conclusion still relevant in the current economic environment? Second, do Divisia economic monetary services provide a better explanation of nominal GDP than simple-sum accounting measures of money? By exploring these questions, the study aims to enhance the understanding of the relevance and effectiveness of Divisia measures in varying economic contexts.

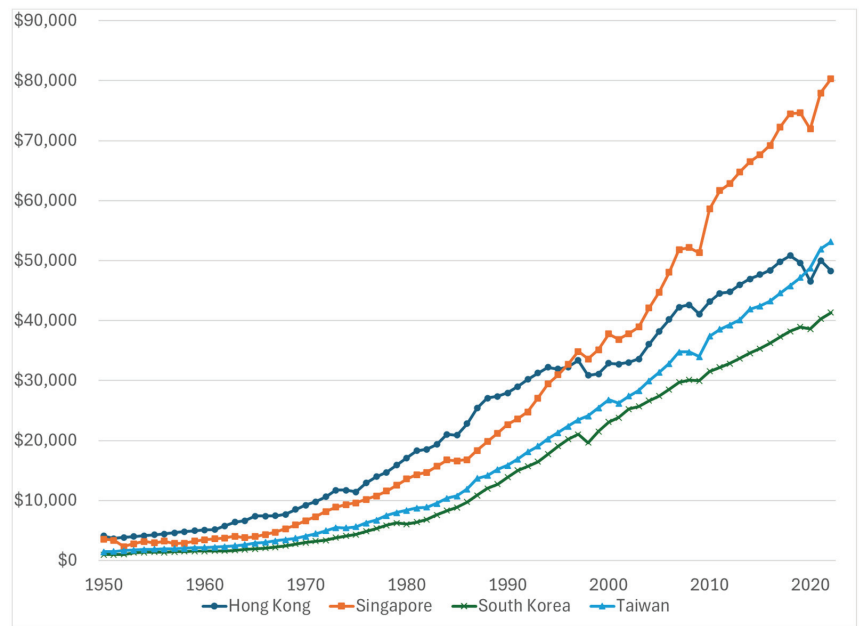


Figure 1. Nominal GDP per capita of the Asian Tigers (1950–2022). Source: Bolt and Van Zanden–Maddison Project Database 2023. Note: These data are adjusted for inflation and differences in the cost of living and are expressed in international dollars at 2011 prices.

This study has three significant contributions and implications. Firstly, it is the only study on economic monetary aggregation that collectively focuses on the Asian Tigers, known for their rapid economic growth. Notably, there has been no research on Korea and Taiwan since 2010, and no studies have examined the Divisia monetary aggregates in Hong Kong. Secondly, this study provides an opportunity to examine whether Divisia monetary aggregates can effectively reflect economic conditions in developed Asian territories. Finally, this study will provide a more accurate financial policy indicator for these Asian territories.

For example, (Barnett et al. 2022b) highlight the importance of the Divisia index for monetary aggregation in Hong Kong, particularly with the emergence of Central Bank Digital Currencies (CBDCs) and the increasing diversity of transaction services. In the case of Korea, the Bank of Korea emphasized the significance of Divisia monetary aggregates in a note (Park and Shim 2018), demonstrating that the Divisia economic index for M2 provides a more accurate reflection of economic growth than the simple-sum accounting monetary aggregate. Furthermore, the Korea Institute of Finance argued that “For quantitative easing (QE) to be successful in the post-global financial crisis era, it is important to measure the actual liquidity that can influence the real economy, which Divisia monetary aggregates can help achieve” (Korea Institute of Finance 2016). In the case of Taiwan, the rapid economic, social, and political changes since the early 1980s led to long-term macroeconomic imbalances. The subsequent financial innovations and deregulation, particularly following the 1989 Banking Act amendment, revealed the limitations of the simple-sum method for measuring money supply, necessitating the adoption of the Divisia index (Binner et al. 2002). Furthermore, Shih (2000) suggests that a re-evaluation of the usefulness of Divisia aggregates as a monetary target variable after the developments of the Taiwanese financial markets reach a certain stage, might be useful. These studies underscore that Divisia monetary aggregates provide a more reliable tool for informing financial policy.

In sum, this study pioneers research on the Asian Tigers jointly, evaluating the effectiveness of Divisia monetary aggregates across various economic contexts and emphasizing their importance as a more accurate financial policy indicator.

We hypothesize that the Divisia monetary index provides a more accurate and robust explanation of nominal GDP fluctuations than the simple-sum method. The Divisia monetary index, by accounting for the different levels of services of various monetary assets, is expected to outperform the traditional simple-sum method in explaining changes to nominal GDP. To test this hypothesis, we comprehensively compare the Divisia economic monetary index with the simple-sum accounting measure across the Asian Tiger territories to assess the effectiveness of the measures in explaining nominal GDP. Specifically, we use the Augmented Dickey–Fuller (ADF) test to check for stationarity, the Johansen cointegration test to analyze long-term relationships, and the Vector Error-Correction Model (VECM) to examine both short-term dynamics and the long-term equilibrium. The objective is to determine whether the Divisia monetary index indeed offers a superior explanation of nominal GDP fluctuations than the simple-sum aggregate.

The remainder of this paper is structured as follows: Section 2 outlines the methodology followed for our analysis, while Section 3 presents the data and identification strategies used for analyzing the issue. Section 4 details the results of the Divisia monetary aggregates and the corresponding test outcomes. Finally, Section 5 discusses the limitations and concludes the study.

2. Methodology

2.1. Simple Sum

Problems with the U.S. monetary demand function using simple-sum money measures were caused by rapid institutional innovation and changes in regulations. Liberalizations of bank management have caused demand for the constituent assets of the available simple-sum monetary accounting aggregates to fluctuate. The mix of monetary services from monetary assets and the savings motive associated with the interest rate payments on monetary assets has varied across components and over time. The opportunity costs of each monetary asset have fluctuated over both time and monetary assets. Measuring the services from aggregated monetary assets by simple-sum aggregation has become inconsistent with aggregation and index number theory and produces the appearance of instability of the demand from the improper measurement of the service flow.

Barnett (1980) identified the fundamental problem in measuring the monetary service flow as an accounting sum. He pointed out that monetary aggregation by accounting summation implies that each component asset is a perfect substitute for each of the others, and he showed how existing economic aggregation and index number theory can solve the problem. In addition, numerous studies have argued that the assumption of perfect substitutability is unreasonable (e.g., Feige and Pearce 1977; Barnett 1982).

Fisher (1922) concluded that the simple sum is the most unreasonable among all the possible indexes he studied in his famous book (also see Barnett et al. 1992). As stated by Fisher (1922, p. 363), the simple sum has the two worst properties of an index number:

“There are two objections to Formula 1, the simple arithmetic (1) that it is simple and (2) that is arithmetic; that it is at once freakish and biased.”

Barnett (1980, 1981) provided the relevant microeconomic aggregation and index number theory to measure monetary services. See (Barnett et al. 1992) for an overview of the relevant derivation and theory.

2.2. Divisia Index

In Divisia aggregation, the opportunity costs of individual monetary asset services are the prices used in computing the share weights of the monetary component growth rates. This opportunity cost is the user cost price of each monetary asset’s services. That price is the interest rate foregone to consume the services of the asset during the period.

The interest foregone at the end of the period is discounted from the end of the period to the beginning of the discrete time period (Barnett 1978). Thus, the opportunity cost is

$$\pi_i = \frac{R - r_i}{1 + R}, \quad (1)$$

where r_i is the expected rate of return on the i th component monetary asset, and R refers to the expected return on the “benchmark” asset, which is an economy’s pure capital providing only store of value and no other services. For example, demand deposits and cash that do not pay interest have the highest user cost price with $r_i = 0$ and, hence, the highest marginal utility. But as in the famous diamonds versus water paradox, high marginal utility does not necessarily mean high total or average utility. While the user cost prices are needed to compute the expenditure shares that weigh the component growth rates, the user cost prices are not weights in the index. They are prices at the margin. In aggregation and index number theory, user cost price aggregates have a dual relationship with monetary quantity aggregates and can be used to produce the implied interest rate aggregates uniquely (Barnett 2010).

Upon deriving the user cost of monetary assets, Barnett (1980) linked the resulting index number theory with microeconomic monetary aggregation theory. Diewert (1976) defined a class of superlative index numbers that bridged the gap between index number theory and aggregation theory. Diewert proved that the Fisher Ideal and Törnqvist–Theil Divisia indices are in a class of superlative indexes providing a second-order approximation to any aggregator function. Barnett correspondingly proposed two quantity indices for monetary aggregation, consisting of the Fisher Ideal and Divisia indices in discrete time with user cost prices. The Törnqvist–Theil discrete time approximation to the Divisia index is the most useful quantity index for policymakers since it has the most easily understood form of any index numbers in Diewert’s (1976) superlative index number class. We shall hereafter refer to the Törnqvist–Theil as just the Divisia index, although, strictly speaking, the Divisia index is the continuous time index attained when discrete time converges to continuous time. Ishida (1984) correspondingly agreed that the Divisia index (in discrete time) is the best statistical index, and he then proceeded to produce Divisia monetary aggregates for Japan.

The Divisia index in continuous time is a line integral approximated during time period t in discrete time by the Törnqvist–Theil index, M_t^D . Let $\prod_i (\cdot)$ designate the multiplication operator over the N component assets, $i = 1, \dots, N$. Then, the resulting discrete-time Divisia index is

$$\frac{M_t^D}{M_{t-1}^D} = \prod_i \left(\frac{x_{i,t}}{x_{i,t-1}} \right)^{\frac{1}{2}(S_{i,t} + S_{i,t-1})} \quad (2)$$

where $x_{i,t}$ is the quantity of good i during period t . In our case, that quantity will be the nominal quantity of monetary asset i , and s_{it} will be asset i ’s share in the user cost evaluated expenditure on the services of all of the N monetary asset components, as follows:

$$s_{it} = \left(\frac{\pi_{it} x_{it}}{\sum_{i=1} \pi_{kt} x_{kt}} \right). \quad (3)$$

Taking the log of each side of Equation (2), we find the equation below and can observe that the growth rate of the Divisia index is a weighted average of the growth rates of each component asset, as follows:

$$\ln M_t^D - \ln M_{t-1}^D = \sum_{i=1} \left[\frac{s_{i,t} + s_{i,t-1}}{2} \right] (\ln x_{i,t} - \ln x_{i,t-1}). \quad (4)$$

The growth rate of the Divisia aggregate can be written as

$$G(M_t^D) = \sum_{i=1}^N s_{it}^* G(x_{i,t}), \quad (5)$$

where $s_{it}^* = 0.5(s_{i,t} + s_{i,t-1})$ and G is the log change operator. Single-period changes, beginning with a base period, could be cumulated to determine the Divisia aggregate level in each period.

2.3. Benchmark Rate

The benchmark rate is an important variable in the user cost prices needed in the computation of the expenditure shares appearing in the Divisia index growth rate. The benchmark rate theoretically reflects the return on a pure capital investment that offers no monetary services, setting an upper limit on the returns on monetary assets that do provide monetary services, as well as a possible lower rate of return. The benchmark rate must never be less than the highest return among the component assets, which provide both monetary services as well as investment returns. In practice, the benchmark rate is often calculated as the highest yield among the assets in the broadest monetary aggregate, or a proxy yield on a high-yielding illiquid asset. The benchmark rate must never be less than the rate of return on any of the monetary component assets to avoid negative user cost prices violating the definition of the benchmark asset.

Selecting an appropriate proxy for the benchmark rate is essential, as it significantly impacts the behavior and accuracy of Divisia monetary aggregates. Countries tailor this concept to their specific contexts. For instance, Hahm and Kim (2000), for Korea, opted for the upper envelope over the yields on 3-year corporate bonds and the yield on all monetary assets, while Barnett and Nguyen (2021) used the prime lending rate in Singapore. In Taiwan, Binner and Kelly (2017) added a 100-basis point liquidity premium to the upper envelope over the rates of return on all monetary assets within the aggregate. The choice of proxy is important to ensure that the user costs accurately reflect the opportunity costs of holding money and align with the financial environment of the respective country. However, there is reasonable robustness of the Divisia index to the choice of benchmark rate, since that rate appears symmetrically in all terms in both the numerator and denominator of the expenditure shares. In contrast, accurate measurement of the rates of return on the component assets is very important, since the share weights are sensitive to those measures.

2.4. Vector Error-Correction Model (VECM)

Based on our results, we evaluate whether Divisia monetary aggregates significantly enhance the explanatory power of changes in nominal GDP (NGDP) compared to simple-sum aggregates. To test this hypothesis, we first verify the order of integration for each of the variables using the Augmented Dickey–Fuller (ADF) test to check for unit root, with the lag lengths determined by the Akaike Information Criterion (AIC) (Akaike 1973). The null hypothesis is that the variable has a unit root. Two variables are cointegrated if there exists a linear combination that is integrated of order zero, $I(0)$, which indicates that the variables have a long-run relationship and share a common stochastic trend over time (Engle and Granger 1987). Cointegration requires that the variables be stationary in their first differences and non-stationary in levels. Based on the findings of the ADF test, we then assess cointegration using the Maximum Likelihood approach proposed by Johansen (1988).

Following these steps, we use the Vector Error-Correction Model (VECM) to analyze whether Divisia monetary aggregates offer a better explanation for NGDP, a widely accepted proxy for economic activity, compared to the simple-sum monetary aggregates. The VECM is particularly suitable for addressing the limitations of the Vector Autoregressive (VAR) model, specifically its inability to account for cointegrated variables that share a long-term equilibrium relationship. This framework allows us to model both short-term

deviations and long-term equilibrium relationships between variables. Previous studies have used similar approaches, employing the ADF test and estimating models with ECM or VECM techniques to assess the explanatory power of Divisia monetary aggregates (Habibullah 1999; Hahm and Kim 2000; Sarwar et al. 2010; Ramachandran et al. 2010; Pua and Hiew 2011; Bhatnagar 2022). The VECM is formulated as follows:

$$\Delta NGDP_t = \beta_{10} + \beta_{11}(L)\Delta NGDP_{t-1} + \beta_{12}(L)\Delta M_{t-1} + \beta_{13}(NGDP_{t-1} - M_{t-1}) + \varepsilon_{1t} \quad (6)$$

$$\Delta M_t = \beta_{20} + \beta_{21}(L)\Delta NGDP_{t-1} + \beta_{22}(L)\Delta M_{t-1} + \beta_{23}(NGDP_{t-1} - M_{t-1}) + \varepsilon_{2t} \quad (7)$$

where L is the lag operator, M_t refers to the respective variable for monetary policy for the four territories (that include simple-sum monetary aggregates, Divisia monetary aggregates, and an interest rate variable), and $NGDP_t$ is the log of quarterly nominal GDP, while $(NGDP_{t-1} - M_{t-1})$ refers to the error-correction term. Following Ramachandran et al. (2010), we test the null hypothesis that $\beta_{13} = 0$ and $\beta_{23} = 0$. If we reject the first and accept the second, then that would imply that the growth rate of output responds to disequilibrium errors in the last period, while the growth rate of money does not.

2.5. Forecast Error-Variance Decomposition (FEVD)

We use a Forecast Error-Variance Decomposition (hereafter referred to as FEVD) analysis to quantify the importance of each shock in explaining the variation in each variable by calculating the fraction of the forecast error variance attributable to each shock at different horizons. This analysis provides insights into the dynamic interactions between variables, revealing the relative importance of various shocks in driving fluctuations.

In the context of monetary policy analysis, FEVD has proven particularly valuable. For instance, when applied to Divisia monetary aggregates, FEVD can demonstrate their superior performance over simple-sum aggregates in explaining variations in economic output. The analysis effectively identifies which monetary aggregates influence output more, underscoring their value in policy analysis. The ability of FEVD to decompose and attribute forecast errors to specific shocks makes it an especially valuable tool for understanding the nuances of monetary policy and economic fluctuations, thereby informing more effective decision-making.

3. Data

3.1. Hong Kong

Hong Kong's financial system is structured around a three-tier banking framework comprising licensed banks, restricted license banks (RLBs), and deposit-taking companies (DTCs). Licensed banks, offering the broadest range of services, significantly influence the narrow money supply (M1) through demand deposits and currency circulation. RLBs principally engage in merchant banking and capital market activities and take deposits of any maturity of HK\$ 500,000 or more. DTCs contribute to broader monetary aggregates, particularly M2 and M3, by accepting deposits of HK\$ 100,000 or more with an original term of maturity of at least 3 months. DTCs also engage in a range of specialized activities, including consumer finance, commercial lending, and securities business.

Significant fluctuations have marked the trajectory of Hong Kong's monetary aggregates following domestic policy changes and external economic pressures. During the 1980s and early 1990s, Hong Kong saw steady growth as it established itself as a global financial center. However, the 1997 Asian Financial Crisis introduced volatility, leading to stringent monetary policies to defend the currency peg and stabilize the financial system. Post-1997, especially after the 2008 global financial crisis, broader monetary measures such as M2 and M3 expanded again from liquidity injections by the Hong Kong Monetary Authority (HKMA) and increased foreign capital inflows.

The monetary aggregates M1, M2, and M3 are primary indicators of the economy's money supply. Appendix B Table A2 provides a basic description of Hong Kong's monetary asset components. The monthly dataset covers January 1999 to June 2024, with the data on levels reported in millions of Hong Kong dollars (HKG). M1, the most liquid form, reflects immediate spending power through currency in circulation and customers' demand deposits. M2 builds upon M1 by including customers' savings and time deposits and negotiable certificates of deposit (NCDs) issued by licensed banks held outside the banking sector, offering a broader perspective on available funds. M3 encompasses all components of M2 plus customers' deposits with restricted license banks and deposit-taking companies, plus NCDs issued by these institutions held outside the banking sector.

The types of liabilities included in the money supply of Hong Kong are similar to those in other economies. However, a major difference relates to the treatment of Government and Exchange Fund placements with banks. The government's deposit holdings are included in the Hong Kong dollar broad money, whereas in most other economies, they are excluded. Furthermore, exchange fund placements with a maturity of over 1 month are counted as customer deposits and, hence, included in the money supply figures.

Although the HKMA reports interest rates for savings and time deposits with licensed banks, the HKMA does not provide the rates for savings and time deposits with restricted licensed banks and deposit-taking companies. We use the average interest rates for time deposits (6 months) and savings deposits of less than HKG 100,000 quoted by the major licensed banks for time and savings deposits included in M2. But for deposits with RLBs and DTCs, we use the 3-month time deposit rate, since DTCs require that deposits have an original term of maturity of at least 3 months. However, due to data limitations, although DTCs require the deposits to be of at least HKG 100,000, and RLBs require the deposits to be of at least HKG 500,000, we use the 3-month deposit rate for deposits of less than HKG 100,000 as a proxy for the rate of deposits with RLBs and DTCs.

In the series for the level of NCDs issued by RLBs and DTCs during the period from May 2011 to April 2015, the quantities reported are zeros, which does not seem credible. We smoothed the data on NCDs for the above period by approximating the zero-quantities by the average of the quantities 12 months before and 12 months after that period. Since the rate of return on NCDs is not reported, we use the yield for 1-year exchange fund bills, which has been included in the broad money aggregates since April 1997. We chose a timeline post-1997 to make use of the "new data" series reported by HKMA. For the benchmark rate, we use the best lending rate quoted by the Hong Kong and Shanghai Banking Corporation Limited and reported by the HKMA.

To compare the performance of Divisia monetary aggregates and simple-sum aggregates in explaining nominal GDP, we use the seasonally adjusted data for nominal GDP between 2009 Q1–2023 Q4 reported by the Census and Statistics Department of the Government of the Hong Kong Special Administrative Region. For the interest rate variable, we use the overnight lending rate provided by HKMA. For our analysis, we use log-transformed variables unless mentioned otherwise. We seasonally adjust all the monetary variables using the X-13 ARIMA (Autoregressive Integrated Moving Average) SEATS procedure.

3.2. South Korea

In Korea, the compilation of various monetary indicators, including the old version of M2, began in March 1981. In 2002, the Bank of Korea revised and published M1 and M2 as new indicators, guided by the IMF's Monetary and Financial Statistics Manual (International Monetary Fund 2000). Since 2006, the Bank of Korea has also prepared and published liquidity indicators Lf and L, resembling M3 and M4 in some other territories. The monetary statistics by economic entities, released in September 2013, show the amount of currency held by each economic entity based on M2 among monetary and liquidity indicators. Until 2002, Korea's monetary and liquidity indicators served as intermediate targets for monetary policy. However, with the adoption of an inflation-targeting system,

the interest rate-oriented monetary policy framework now uses interest rates as informative variables for policy management.

Appendix B Table A3 provides a basic description of Korea's monetary asset components, covering a monthly dataset from January 2009 to June 2024, with the data on levels reported in millions of Korean won (KRW). The data, sourced from the Economic Statistics System (ECOS) of the Bank of Korea, are based on monthly averages and have been seasonally adjusted. The Bank of Korea classifies the primary components of M1 (narrow money) as currency in circulation, demand deposits, and transferable savings deposits. M2, which includes M1, refers to broad money, encompassing money market funds (MMFs), short-term time and savings deposits, beneficiary certificates, marketable financial instruments, short-term financial debentures, short-term money in trust, and others. The maturity of all assets in M2 is within 2 years. Lf includes liquid financial instruments. L includes long-term assets (over 2 years), such as long-term financial instruments and life insurance reserves.

Unfortunately, the Bank of Korea does not report matched interest rates for all assets¹. When the Bank of Korea collects data, it gathers total amounts and interest rates separately from various institutions, which can lead to mismatches between the two². For instance, short-term financial debentures and short-term money in trust do not have corresponding interest rates. In such cases, to best reflect the interest rates of short-term assets, the most relevant proxy interest rates were used, such as the interest rates of 1-year treasury bonds and 91-day CDs, which best capture the characteristics of short-term assets. For Lf's life insurance reserves, the 3-year treasury bond rate was used as a proxy for the interest rate on life insurance because treasury bond yields are directly employed in insurance product design, pricing, and reserve calculations, and they also serve as a key factor influencing the overall financial performance of insurance companies (Lee 2016)³. L includes long-term assets such as liquid financial instruments like government and municipal bonds. In this case, to adjust for the yield curve, the rate of return on long-term assets was replaced with the interest rate on comparable short-term assets, specifically the 1-year Treasury bond rate.

This paper will use the corporate bond rate (BBB-) as the benchmark rate for Korea, following the approach of Hahm and Kim (2000) for compatibility with their results. However, there are two more reasons. First, there are data limitations for Korea. Barnett et al. (2013) measured the benchmark rate as the weighted average effective loan rate for low-risk, 31–365 days' loans across all commercial banks from the FRED database maintained by the St. Louis Federal Reserve Bank. A comparable loan rate does exist in Korea and typically exceeds deposit interest rates. However, the loan rate is a weighted average rate for all loans, not specifically for low-risk short-term loans. Therefore, in the case of Korea, it is challenging to distinguish between various risk levels and loan periods. Secondly, corporate bonds in Korea are categorized by ratings such as AA- and BBB-, which allows for a more precise differentiation of risk levels. Moreover, the Bank of Korea provides data on corporate bonds with a 3-year maturity, which will be used in this study. While a 3-year bond rate is relevant to the theory, we do not use long-term bond rates, which would be too far removed from the necessary short-term rate of return for an accurate benchmark.

To compare the performance of Divisia monetary aggregates and simple-sum aggregates in explaining nominal GDP, we use the seasonally adjusted data for nominal GDP between 1999 Q1 and 2024 Q1 reported by the Federal Reserve Bank of St. Louis. For the interest rate variable, we use the overnight lending rate provided by the Bank of Korea. For our analysis, we use log-transformed variables unless mentioned otherwise.

3.3. Singapore

Singapore's monetary policy framework employs an institution-based approach to measuring money supply, aligning with methodologies used in the United Kingdom and Germany. This approach defines monetary aggregates based on the liabilities of specific financial institutions rather than the characteristics of monetary instruments. The narrowest

measure, M1, includes currency in active circulation and private sector demand deposits within the banking system. M2 expands upon M1 by incorporating quasi-money, while M3, the broadest measure, further includes net deposits with non-bank financial institutions (NBFIs). The banking system comprises the Board of Commissioners of Currency, Singapore (BCCS), and commercial banks, forming the foundation for M1 and M2 calculations. M3 extends this scope by integrating net deposits from finance companies, offering a more comprehensive view of the monetary landscape.

A significant modification to this framework, which has evolved over time, occurred in November 1998, following the Post Office Savings Bank's (POSB) merger with the Development Bank of Singapore (DBS). This consolidation required a recalibration of monetary measurements, as POSB's deposits were reclassified on par with other commercial banks and thus incorporated into M1 and M2 aggregates. Additionally, the computation of M3 was adjusted to include POSB's term deposits held with the Monetary Authority of Singapore (MAS), which were previously excluded. To maintain consistency in monetary data, authorities implemented retrospective revisions dating back to October 1982, when POSB initially placed deposits with MAS. This restructuring underscores Singapore's commitment to precise monetary statistics and adapting to evolving institutional dynamics. Consequently, Singapore's monetary policy framework is anchored in a tripartite system of monetary aggregates—M1, M2, and M3—each serving as a distinct indicator of money supply with progressively expanding scope and complexity.

For Singapore, we use monthly data from January 1991 to June 2021, with data reported in millions of Singapore dollars (SGD) from the MAS. The period concludes in June 2021 due to the transition of the commercial bank interest rate data from monthly to quarterly reporting. The MAS reports three levels of simple-sum monetary aggregates. Appendix B Table A4 provides a basic description of Singapore's monetary asset data. M1 includes currency in circulation and demand deposits in banks, both carrying zero interest rates. M2 expands upon M1 by adding fixed deposits, savings deposits, and negotiable certificates of deposit (NCDs) within the banking sector. M3 further broadens the definition of money by including net deposits in finance companies. According to MAS guidelines, POSB's fixed deposits and savings deposits have been included within the respective commercial banks, while the remaining components of POSB have been incorporated into M1. Here, "banks" refers to commercial banks, and "finance" refers to finance companies.

Barnett and Nguyen (2021) previously examined Singapore's financial data and proposed alternatives to address data limitations. Our paper not only refines their approach but also underscores the significant role of their proposals. To achieve a more accurate composition of monetary aggregates, POSB's non-fixed deposits and non-savings components were included in M1, which aligns with Singapore's banking guidelines. Following DBS's acquisition of POSB in November 1998, POSB's data were incorporated into M1 and M2, with separate data no longer provided. Consequently, M2 now includes POSB's fixed deposits and savings, while M1 comprises the remaining components. The allocation of POSB data follows Barnett and Nguyen's (2021) methodology, assuming POSB interest rates align with those of the banking sector.

For M3, net deposits with finance companies were divided into fixed and savings deposits using the net deposits' liabilities ratio and corresponding 3-month interest rates. Due to the unavailability of the NCDS interest rate for M2, we substitute the 3-month T-bill rate. Barnett and Nguyen (2021) used the 3-month commercial bill rate but switched to the T-bill rate post-2013 after the MAS ceased providing the former data. The 3-month T-bill rate is chosen as it is a short-term, low-risk asset with comprehensive data available.

This paper adopts the short-term lending rate as the benchmark rate for Singapore, following the methodology of Barnett et al. (2013), for three principal reasons. First, from a theoretical perspective, the benchmark rate should reflect the rate of return on pure capital and must exceed the returns on all monetary assets that provide services to depositors. Second, from an economic standpoint, it is unrealistic for banks to offer deposit rates

that exceed their investment yields. Lastly, in the specific context of Singapore, the prime lending rate consistently exceeds the interest rates on the component monetary assets.

To compare the performance of Divisia monetary aggregates and simple-sum aggregates in explaining nominal GDP, we use the seasonally adjusted data for nominal GDP between 2005 Q4 and 2020 Q2 as reported by the Singapore Department of Statistics. For the interest rate variable, we use the overnight lending rate provided by the Monetary Authority of Singapore. Due to the interest rates being reported from 2005 to 2020, we choose that timeline. For our analysis, we use log-transformed variables unless mentioned otherwise.

3.4. Taiwan

Taiwan's financial system, while sharing similarities with other advanced economies in East Asia, possesses distinct characteristics that set it apart. The system is characterized by a well-developed banking sector and a robust regulatory framework. The Central Bank of the Republic of China (Taiwan), the nation's central bank, plays a pivotal role in regulating monetary aggregates. These aggregates are essential indicators of the economy's money supply and financial stability, and the central bank's management of them is a key aspect of the financial system's stability.

A unique feature of Taiwan's financial landscape is the role of Chunghwa Post, the official postal service. Beyond its traditional functions, Chunghwa Post operates a postal savings system, forming a significant component of the nation's broader financial infrastructure. This system allows individuals to hold savings accounts and make deposits, contributing substantially to the overall monetary aggregates, particularly M2. Integrating postal services with financial functions provides Taiwan with a distinctive mechanism for mobilizing domestic savings and enhancing financial system stability.

Over time, Taiwan's monetary aggregates have shown significant changes in response to both internal and external economic factors. The rapid industrialization of the 1970s and 1980s led to substantial increases in M1 and M2, followed by a period of stabilization in the 1990s as the economy matured. Global events, such as the 1997 Asian Financial Crisis and the 2008 Global Financial Crisis, prompted shifts in monetary strategy. The central bank responded by implementing various measures to manage liquidity and ensure financial stability. In the post-2008 era, the central bank has maintained a cautious approach, balancing economic stimulus with concerns about asset bubbles and inflation, resulting in moderate growth in monetary aggregates with occasional adjustments.

Appendix B Table A5 provides a basic description of Taiwan's monetary asset components. The monthly dataset covers January 2005 to June 2024 with the data on levels presented in millions of New Taiwan dollars (TWD). In terms of monetary aggregates in Taiwan, the central bank distinguishes between M1A and M1B, with M1A representing currency in circulation, checking accounts, and passbook deposits, while M1B includes M1A plus passbook savings deposits. M2 expands on M1B by including savings deposits, time deposits, foreign currency deposits, postal savings deposits, negotiable certificates of deposit (NCDs), money market funds (MMFs), repurchase agreements, and government deposits.

The data on levels for MMFs provided by the central bank start from October 2004. Therefore, for our analysis, we use the data starting from January 2005 to June 2024. However, the data for MMFs have not been reported since May 2017. We use a regression of the series to fill in the missing data.

For the rates of return on passbook deposits, passbook savings deposits, time and savings deposits, and foreign currency deposits of commercial banks, we use the respective interest rates quoted by the Bank of Taiwan and reported by the central bank. For the rate of return on NCDs, we use the 91–180 days' NCDs rate in the secondary market. For the rates of return on repurchase agreements, we use the 91–180 days' repos rate on government bonds in the secondary market. Since the rates of return on MMFs are not reported, we use the overnight interbank loan rate. For the benchmark rate, we use the average base lending

rate offered by the five major banks as published by the Central Bank of the Republic of China (Taiwan).

To compare the performance of Divisia monetary aggregates and simple-sum aggregates in explaining nominal GDP, we use the seasonally adjusted data for nominal GDP between 2005 Q1–2024 Q1 reported by the National Statistics, Republic of China (Taiwan). For the interest rate variable, we use the overnight lending rate provided by the Central Bank of the Republic of China (Taiwan). For our analysis, we use log-transformed variables unless mentioned otherwise. We seasonally adjust all the monetary variables using the X-13 ARIMA (Autoregressive Integrated Moving Average) SEATS procedure.

4. Results

4.1. Divisia Monetary Aggregates Results

4.1.1. Hong Kong

In Hong Kong, as shown in Figures 2a and 3a, M1 and DM1 are identical, as is expected since they consist of the same components with zero interest rates, leading to identical user cost prices. However, a significant divergence exists between the narrower M1/DM1 and the broader M2 and M3 measures for both simple-sum and Divisia aggregates. Simple-sum measures of M2 and M3 are generally higher than their Divisia counterparts, suggesting that simple-sum measures overstate the quantity of monetary services in the economy.

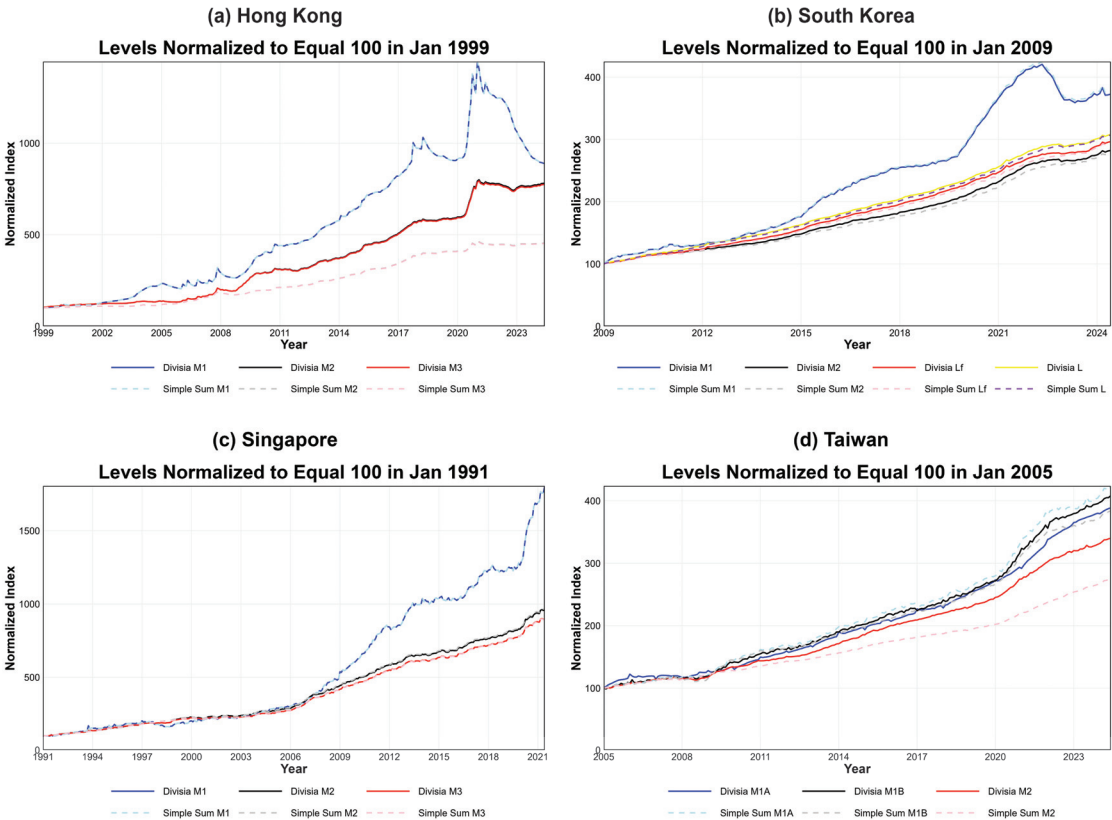


Figure 2. Divisia versus simple-sum monetary aggregates for the four Asian tigers.

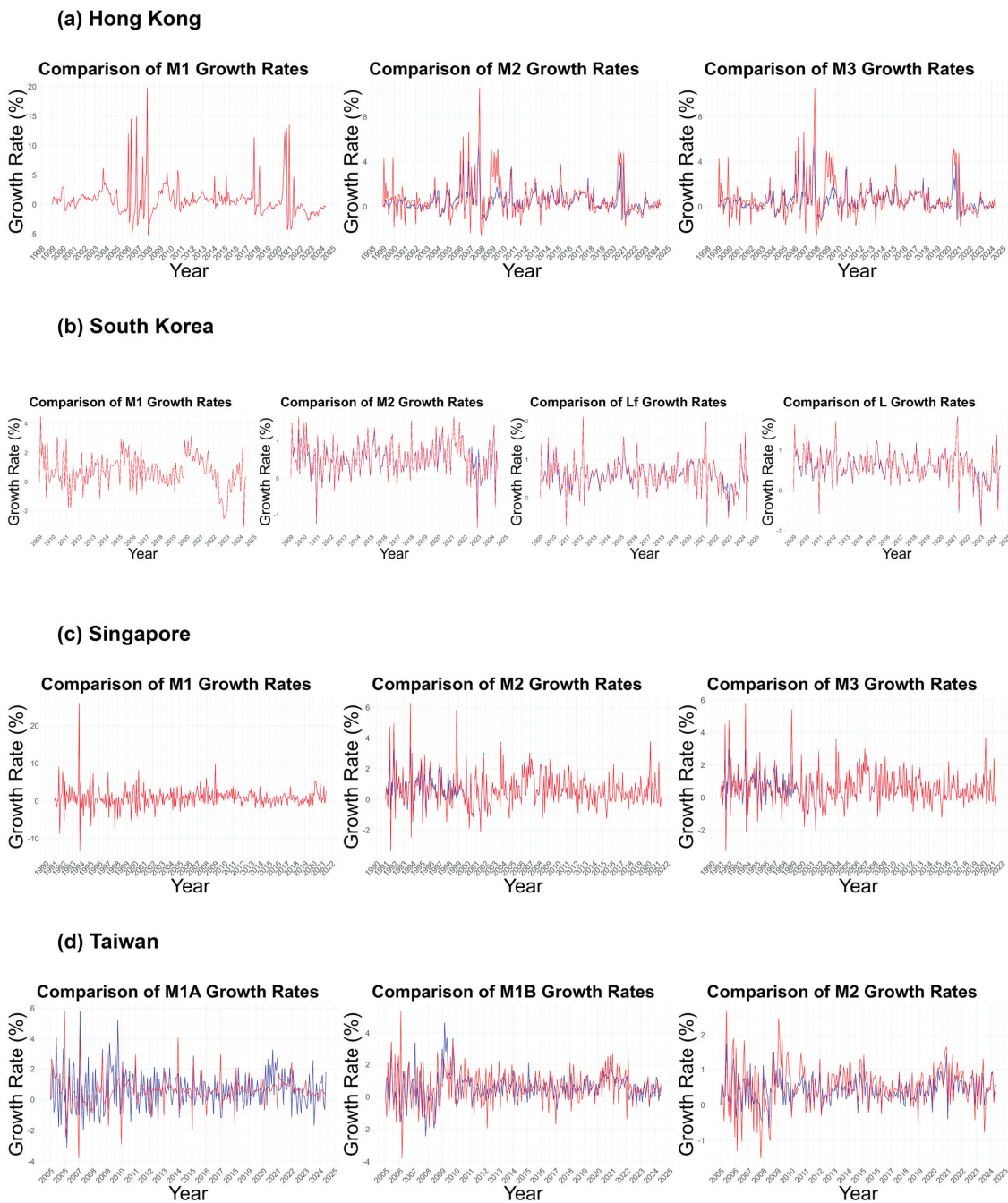


Figure 3. Divisia versus simple-sum growth rates for the four Asian tigers.

Over time, particularly after 2008, the gaps between simple-sum and Divisia measures widen, likely due to financial innovations and interest rate changes. Divisia measures capture fluctuations, such as those during the global financial crisis and the COVID-19 pandemic, more distinctly, thereby highlighting their sensitivity to changes in the monetary

environment. This underscores the importance of using Divisia aggregates for a more accurate understanding of monetary conditions in Hong Kong and the crucial role they play in shaping effective monetary policy.

4.1.2. South Korea

In Korea, as depicted in Figures 2b and 3b, M1 and DM1 are almost identical, reflecting that M1's components, including cash and demand deposits, have zero or very low interest rates. While there are slight differences between simple-sum M2 and DM2, with DM2 generally being slightly higher, the difference between them remains consistent over time.

A more noticeable difference is seen between Lf and DLf, with DLf being consistently higher. Additionally, DL follows a similar trend, maintaining a higher level than the corresponding simple-sum measures. Clearly, the broad simple-sum monetary aggregates undervalue the growth rate of monetary services produced by its components.

All monetary aggregates show an upward trend over time, with a notable increase after 2015. The consistent elevation of Divisia indices above simple-sum measures indicates that in Korea, the simple-sum measures undervalue the growth of monetary services in the economy, reflecting the structure of Korea's financial market, interest rate environment, and financial instrument substitutability. Over time, the widening divergence between Divisia and simple-sum indices underscores the importance of using Divisia aggregates for accurate monetary policy analysis in Korea.

4.1.3. Singapore

In Singapore, Figures 2c and 3c reveal notable differences between simple-sum and Divisia measures, especially in the broader aggregates such as M2 and M3. M1 and DM1 are nearly identical in Singapore, as the components of M1, such as currency and demand deposits, typically have low or zero interest rates, leading to equivalent user cost prices.

However, unlike in Korea where Divisia measures were consistently higher, Singapore's DM2 and DM3 are slightly lower than their simple-sum counterparts, with nearly identical growth rates. This aligns with findings from Barnett and Nguyen (2021), that Divisia measures behave most differently from simple-sum measures when interest rates are high. A significant structural change in November 1998, when simple-sum M2 peaked due to the acquisition of POSB into the banking sector, is not reflected in Divisia's measures, indicating their robustness to institutional changes not altering demand for monetary services.

During the 1997–1998 Asian financial crisis, a notable contraction in money services was observed in the growth rates of DM2 and DM3, while this change was less apparent in M2 and M3, thereby indicating that the Divisia index is more sensitive to economic shocks than the simple sum accounting stock.

4.1.4. Taiwan

In Taiwan, the analysis of monetary indicators, as shown in Figures 2d and 3d, reveals significant differences between M1A and DM1A. Unlike other territories, Taiwan's M1A includes passbook deposits that typically carry higher interest rates. In contrast, DM1A and DM2 show levels lower than their simple sum counterparts.

Over time, all monetary indicators exhibit an upward trend, although growth rates fluctuate during major economic events such as the 2008 global financial crisis and the 2020 COVID-19 pandemic. Divisia indices appear more sensitive to these shocks, and there is a trend of widening gaps between Divisia and simple-sum indices over time, reflecting the impact of financial innovations and interest rate changes.

The Divisia indices' ability to capture changes in economic conditions more accurately underscores their importance in Taiwan's monetary policy decisions. The differences between Divisia and simple-sum indices in M1B and M2 can be attributed to the structural characteristics of Taiwan's financial market and the substitutability of various financial products. The greater sensitivity of Divisia indices to economic shocks and financial innovations suggests they provide a more accurate reflection of Taiwan's monetary environment.

The findings of this analysis significantly differ from the earlier paper by Shih (2000), as expected from the rapid development of Taiwan’s financial markets since the late 90s.

In sum, the analysis across Hong Kong, Korea, Singapore, and Taiwan demonstrates that Divisia monetary aggregates offer a more precise reflection of economic conditions than simple-sum measures, particularly during financial innovations and economic shocks. This is attributed to the Divisia indices’ heightened sensitivity to interest rate changes and their superior ability to capture the liquidity and substitutability of financial assets, underscoring their importance for accurate and effective monetary policy analysis in these regions.

4.2. VECM Results

4.2.1. Hong Kong

Appendix B Table A6 shows the results of the ADF test for all the variables in the log level and the first difference. The results show the presence of a unit root in the log level for all variables but no presence of a unit root in the first differences, thereby indicating that the order of integration of the variables is 1.

We then carry out the Johansen cointegration test using $p - 1$ lags, where p is determined by the Akaike Information Criterion (AIC). Based on the results, as presented in Appendix B Table A7, where “r” represents the cointegration rank, we conclude that at a 5% level of significance M1 and DM1 are cointegrated with NGDP. On the other hand, M2, DM2, M3, DM3, and R are not cointegrated with NGDP. We then proceed to fit a VEC model for M1 and DM1. Table 1 summarizes the results and Table 2 shows the results of the cointegration space. We note that in the cases of both M1 and DM1, we reject the null hypothesis that $\beta_{13} = 0$ and accept the null hypothesis that $\beta_{23} = 0$. This implies that the growth rate of both simple-sum and Divisia money are exogenous to NGDP and are, therefore, predictors of NGDP. In the case of Hong Kong, both simple-sum and Divisia aggregates perform equally well in terms of predicting NGDP. In Hong Kong, narrow money contributes towards the prediction of NGDP more significantly than broad money.

Table 1. Error-Correction Term.

Variables	ECT in Output Equation	ECT in Money Equation
M1	Significant	Not significant
DM1	Significant	Not significant
M2 ¹	--	--
DM2 ¹	--	--
M3 ¹	--	--
DM3 ¹	--	--
R ¹	--	--

¹ No cointegrating relationship exists between GDP and M2, DM2, M3, DM3, and R, respectively.

Table 2. Results of the cointegration space.

NGDP and M1		
Eigenvectors, normalized to the first column:		
Variables	NGDP	M1
NGDP	1.000	1.000
M1	−0.398 ***	−0.134
Loading matrix		
Variables	NGDP	M1
ΔNGDP	−0.231 ***	−0.00314 ***
ΔM1	−0.0262	−0.0505

Table 2. Cont.

NGDP and DM1		
Eigenvectors, normalized to the first column:		
Variables	NGDP	DM1
NGDP	1.000	1.000
DM1	−0.398 ***	−0.134
Loading matrix		
Variables	NGDP	DM1
ΔNGDP	−0.231 ***	−0.00314 ***
ΔDM1	−0.0262	−0.0505

Note: Significance codes: *** 0.01, ** 0.05, * 0.1.

4.2.2. South Korea

Appendix B Table A8 shows the ADF test results for all the variables at the log level and in the first difference. The ADF test results for Korea show that although GDP, M2, Lf, L, DM2, DLf, DL, and R are integrated to the order of 1, M1 and DM1 are not. They are in fact integrated in the order of 2. Because candidates for inclusion in a cointegrating vector must share an identical order of integration, we do not perform the Johansen cointegration tests for GDP and M1 and GDP and DM1. However, conducting the Johansen cointegration tests between GDP and the rest of the variables (M2, Lf, L, DM2, DLf, and DL) presented in Appendix B Table A9 shows no presence of cointegrating relationships.

4.2.3. Singapore

Appendix B Table A10 shows the ADF test results for all the variables at the log level and in the first difference for Singapore. The test results show that although M1, M2, M3, DM1, DM2, DM3, and R are integrated to the order of 1, NGDP is not. Because candidates for inclusion in a cointegrating vector must share an identical order of integration, we do not perform the Johansen cointegration tests for Singapore.

4.2.4. Taiwan

Appendix B Table A11 shows the results of the ADF test for all the variables at the log level and in the first difference. The results show the presence of a unit root in the log level for all variables but no presence of a unit root in the first difference, thereby indicating that the order of integration of the variables is 1.

We then conduct the Johansen cointegration test using $p - 1$ lags where p is determined by the Akaike Information Criterion (AIC). Based on the result, as presented in Appendix B Table A12, we conclude that at a 5% level of significance, DM1A is cointegrated with NGDP. On the other hand, M1A, M1B, M2, DM1B, DM2, and R are not cointegrated with NGDP. We then proceed to fit a VEC model for DM1A. Table 3 below summarizes the results, and Table 4 shows the results of the cointegration space. We note that in the case of DM1A, we reject the null hypothesis that $\beta_{13} = 0$ and accept the null hypothesis that $\beta_{23} = 0$. This implies that the growth rate of Divisia money supply is exogenous to NGDP and can, therefore, contribute to predicting NGDP. Similar to Hong Kong, in Taiwan, Divisia narrow money contributes more significantly towards the prediction of NGDP as opposed to broad money.

Table 3. Error-Correction Term.

Variables	ECT in Output Equation	ECT in Money Equation
M1A ¹	--	--
DM1A	Significant	Not significant
M1B ¹	--	--
DM1B ¹	--	--
M2 ¹	--	--
DM2 ¹	--	--
R ¹	--	--

¹ No cointegrating relationship exists between GDP and M1A, M1B, DM1B, M2, DM2 and R, respectively.

Table 4. Results of the cointegration space.

NGDP and DM1A		
Eigenvectors, normalized to the first column:		
Variables	NGDP	DM1A
NGDP	1.000	1.000
DM1A	−0.554 ***	−0.320
Loading matrix		
Variables	NGDP	DM1A
ΔNGDP	−0.386 ***	−0.00420 ***
ΔDM1A	−0.0117	0.0170

Note: Significance codes: *** 0.01, ** 0.05, * 0.1.

4.3. FEVD Results

4.3.1. Hong Kong

Figures 4 and 5 depict the variance decomposition for the models with M1 and DM1 in Hong Kong. The FEVD analysis confirms our findings above. We see that both M1 and DM1 are almost completely exogenous in the sense that almost none of their variances are being explained by shocks in NGDP. On the other hand, shocks in M1 and DM1 explain a significant fraction of the variances in NGDP. By the 10th quarter, those shocks explain almost 40% of the fluctuations in NGDP. It is important to note that, since the interest rates are zero for the components in M1, the simple sum and Divisia measures for narrow money in Hong Kong are the same. Therefore, the results of the FEVD analysis for M1 and DM1 are identical.

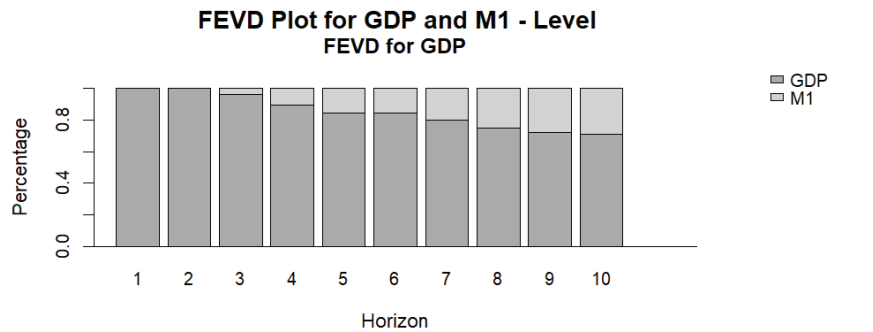


Figure 4. Cont.

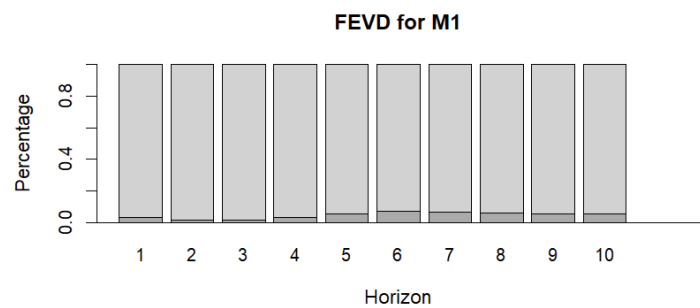


Figure 4. FEVD Plot for GDP and M1.

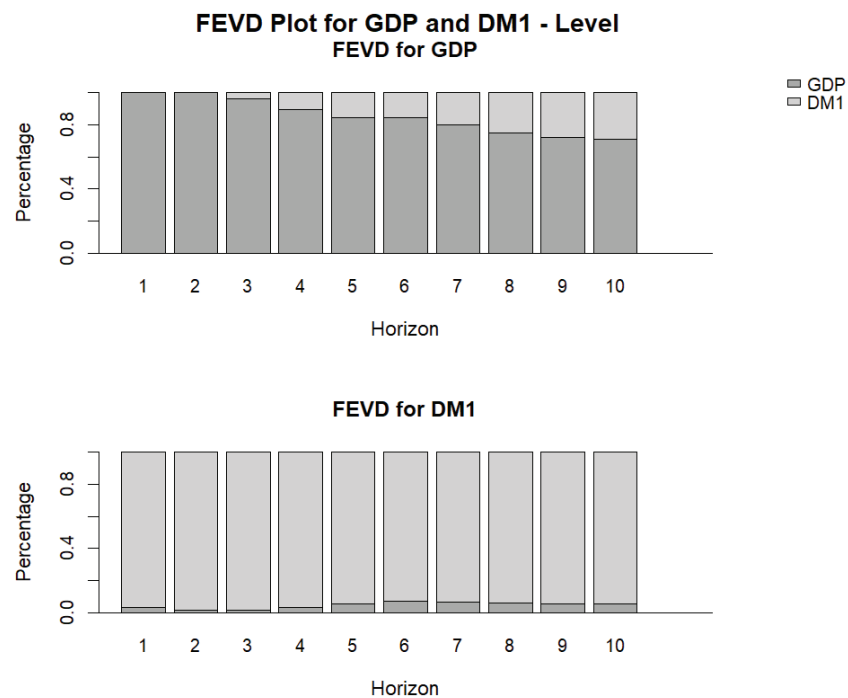


Figure 5. FEVD Plot for GDP and DM1.

4.3.2. Taiwan

Figure 6 depicts the variance decomposition for the model with DM1A in Taiwan. We can see that DM1A is completely exogenous in the sense that almost none of its variance is explained by shocks in NGDP. On the other hand, shocks in DM1A explain a significant fraction of the variances in NGDP. By the 10th quarter, those shocks explain almost 40% of the fluctuations in NGDP.

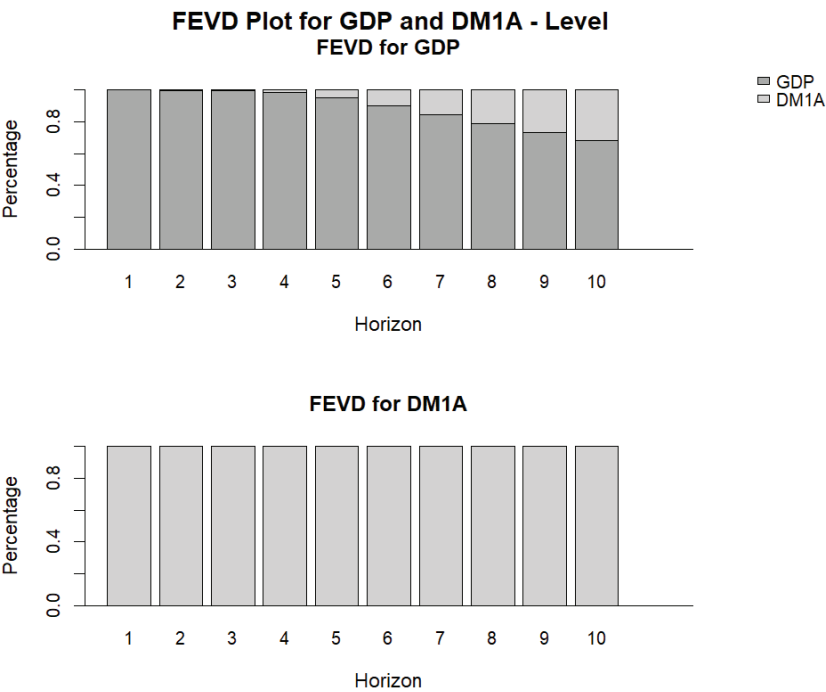


Figure 6. FEVD Plot for GDP and DM1A.

4.4. Forecast Evaluation and Residual Analysis

4.4.1. Hong Kong

To check for the reliability of the Vector Error Correction model (VECM) used in Section 4.2, we perform a dynamic factor model and time-series analysis of the residuals. We backtest the VECM by producing Monte Carlo simulations with 1000 replications to forecast 10 horizons. We use the forecast performance metrics of mean absolute error (MAE), mean error (ME), and root mean square error (RMSE) for 10 horizons, shown in Appendix B Figure A1, to evaluate the model’s forecast performance. The MAE, ME, and RMSE all increase with the forecast horizon for both GDP and M1, with M1 consistently showing higher error rates. This suggests that the model’s predictions for money supply are less accurate than for GDP. The positive and increasing ME for both variables indicates a systematic overestimation in the model’s forecasts. Overall, the forecast performance for Hong Kong suggests that the VECM model’s accuracy diminishes for longer-term predictions, with money supply being particularly challenging to forecast accurately.

We then perform dynamic factor analysis as well as residual analysis on the VECM as shown in Appendix B Table A13. The factor loadings show that both GDP and M1 have equal weights on the first factor (PC1), while they have opposite weights on the second factor (PC2). This suggests that PC1 captures overall economic activity, while PC2 might represent divergences between GDP and money supply. The correlations between VECM residuals and factors are relatively low, indicating that the factors capture most of the systematic variation in the data. Residual diagnostics reveal normality, no significant correlation, and no volatility clustering. The analysis of GDP and DM1 are similar. Since the interest rates are zero for the components in M1, the simple sum and Divisia measures for narrow money in Hong Kong are the same.

4.4.2. Taiwan

Similarly, for Taiwan, we also present the forecast performance metrics of mean absolute error (MAE), mean error (ME), and root mean square error (RMSE) for 10 horizons, as shown in Appendix B Figure A2, to evaluate the model's forecast performance. Taiwan's VEC model demonstrates more stability and generally lower forecast errors compared to Hong Kong. The MAE initially increases but then slightly decreases for longer horizons, with GDP showing a higher MAE than DM1A. The ME is negative for both variables, indicating a tendency for the model to underestimate, though the bias for DM1A approaches zero at longer horizons. The RMSE increases with the forecast horizon, with GDP and DM1A showing similar patterns.

We then perform dynamic factor analysis as well as residual analysis on the VECM as shown in Appendix B Table A14. The factor loadings show that both GDP and DM1A have equal weights on the first factor (PC1), while they have opposite weights on the second factor (PC2). This suggests that PC1 captures overall economic activity, while PC2 might represent divergences between GDP and money supply. The correlations between VECM residuals and factors are relatively low except for that between the residuals of GDP and Factor 2, indicating that the factors capture most of the systematic variation in the data. Residual diagnostics reveal near normality, no significant correlation, and no volatility clustering.

5. Conclusions

A large number of theoretical studies as well as empirical studies have repeatedly shown that Divisia monetary aggregates are superior to their simple sum counterparts, which have no competent foundations in economic theory. Nevertheless, many central banks in the world, including those of developed economies, such as the Hong Kong Monetary Authority (HKMA), Bank of Korea (BOK), Monetary Authority of Singapore (MAS), and the Central Bank of the Republic of China (Taiwan), continue reporting money supply as a simple sum. This paper provides the construction of Divisia monetary aggregates jointly for the Asian Tigers, the four developed Asian economies, Hong Kong, South Korea, Singapore, and Taiwan, and explores the link between various monetary aggregates, interest rates, and output. Our analysis demonstrates that for developed Asian economies, Divisia monetary aggregates offer a more precise reflection of economic conditions than simple sum measures, particularly during financial innovations and economic shocks. Our analysis also suggests that for both Hong Kong and Taiwan, growth rates of Divisia monetary aggregates contribute to predicting nominal GDP. However, contrary to evidence in other advanced economies, such as the US, narrow money explains a larger fraction of the variance in economic activity than broad money.

Narrow money tends to respond more sensitively to short-term fluctuations in economic activity because narrow money is more directly and rapidly influenced by monetary policy in our sample. In contrast, broad money includes deposits and financial assets, which may be less sensitive to short run fluctuations, especially in economies where economic agents hold financial assets for longer-term purposes. For example, Friedman and Schwartz (1963) demonstrate how changes in the money supply impact economic activity differently over time, emphasizing the more immediate effects of narrow money. The higher explanatory power of narrow money in this sample may be due to monetary policy having a more pronounced and rapid effect on short-term cash and liquid assets, compared to financial assets held for longer durations in the Asian Tiger economies, perhaps for cultural reasons and slow subjective rates of time discount.

A significant portion of our analysis depended on data reported by the respective central banks of Hong Kong, South Korea, Singapore, and Taiwan. However, due to various limitations in data availability, we had to choose relevant proxy variables for certain interest rates, since the exact rates for the relevant variables included in the money supply of these territories were not reported by the respective central banks.

Author Contributions: All three authors contributed in major ways to this research. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: All governmental sources clearly identified in the paper.

Acknowledgments: When relevant, acknowledgements mentioned in footnotes. We are truly grateful for the valuable feedback from the two anonymous reviewers.

Conflicts of Interest: The authors declare no conflicts of interest. All errors are our own.

Appendix A

Table A1. Recent Country-specific Studies on Divisia Monetary Aggregates.

Authors	Country	Data Sources	Components	Years	Research Design	Main Results
Anderson and Jones (2011)	US	Federal Reserve Board, Bank Rate Monitor Corporation, FRED	M1, M2	1967–2011	Divisia monetary aggregates, Revision of Monetary Services Indexes	Monetary services indexes (MSI) using Divisia monetary aggregation improves the accuracy of economic forecasts compared to simple sum aggregates, especially for broader monetary measures.
Barnett et al. (2022b)	China	People’s Bank of China, China Statistical Yearbook, Ant Group IPO prospectus, Alibaba China.	M0, M1, M2, M3	2000–2020	Divisia monetary aggregates, Spectral analysis, Dickey–Fuller Test	Divisia aggregates outperform simple sums, especially for broader aggregates. Augmented Divisia captures COVID-19 effects. All aggregates have low short-run but high long-run coherence with GDP. Monetary aggregates lag GDP in the short run.
Barnett et al. (2016)	India	OECD database, Index Mundi, Econ Stats website, (Ramachandran et al. 2010)	M1, M2, M3	2000–2008	SVAR analysis, Impulse response analysis, Variance Decomposition, Flip-Flop analysis, Forecasting	Models with Divisia monetary aggregates outperform those with simple sum aggregates in explaining exchange rate fluctuations and monetary policy impacts.
Belongia and Ireland (2022)	Euro Area	ECB, Bruegel	M1, M2, M3	2001–2019	P-star regressions for the price level	Divisia monetary aggregates significantly predict inflation, with larger and statistically significant coefficients when the sample includes the COVID-19 pandemic period.
Belongia and Ireland (2021)	UK	Bank of England	Not applicable	1978–2019	P-star model, H-P filter	Divisia monetary aggregates provide a more stable framework for nominal income targeting under the zero lower bound.
Belongia and Ireland (2016)	US	Center for Financial Stability, Federal Reserve	M1, M2	1967–2013	Structural VAR	Strong correlations between money supply measures and economic output/prices, especially when using Divisia monetary aggregates instead of simple-sum measures, suggest monetary policy plays an important role in business cycles.

Table A1. *Cont.*

Authors	Country	Data Sources	Components	Years	Research Design	Main Results
Benchimol (2016)	Israel	Bank of Israel, Michelson and Suhoi (2014)	M1, M2	1995–2013	New Keynesian DSGE models, Bayesian estimations, Simulations	Divisia monetary aggregates significantly impact output during crises, offer better forecasting accuracy, and money shocks predict financial risks.
Berar and Owladi (2013)	UK	Bank of England, Bankstats	Not applicable	2011–2012	Description of amendments to Divisia money series calculation	Switching the ISA interest rate in Divisia calculations from quoted to effective rates leads to downward revisions in Divisia money growth series, especially for the household sector.
Bhatnagar (2022)	India	Reserve Bank of India, Central Statistics Office	M1, M2, M3	1999–2019	VECM, FEVD	Divisia aggregates with narrow money better predict economic activity compared to simple sum aggregates.
Binner and Kelly (2017)	Taiwan	Central Bank of the Republic of China (Taiwan), DataStream	M2	1985–2016	Block recursive structural VAR model	Divisia monetary aggregates resolve short-run price, output, and exchange rate puzzles and provide sensible long-run impulse responses to monetary shocks.
Bissoondeal et al. (2010)	UK	Bank of England Datastream	M1, M2	1977–2008	Cointegration analysis, Maximum likelihood approach of Johansen and Juselius (1990) methods, Weak exogeneity tests, Parameter constancy tests	Long-run money demand relationships exist for UK Divisia and simple sum aggregates, with similar elasticities and minor differences in share price effects.
Brill et al. (2021)	Euro Area	ECB Statistical Data Warehouse, MFI	M1, M2	2003–2018	Divisia monetary aggregates, Panel probit analysis	Divisia monetary aggregates for the euro area predict recessions more accurately than simple-sum aggregates and are useful for macroeconomic analysis and forecasting of country-specific monetary conditions.
Chen and Nautz (2015)	Germany	Deutsche Bundesbank, ECB, Eurostat, MFI	M3	2001–2014	Divisia monetary aggregates, Structural VAR models, DM-test	The divergence between M3 and Divisia M3 provides superior forecasts for German output growth during the Great Recession compared to models using only M3 or Divisia M3.
Darvas (2015)	Euro Area	ECB, Bruegel, Eurostat	M1, M2, M3	2001–2014	Structural VAR models, Impulse response functions	Divisia monetary aggregates significantly impact output and prices, while simple-sum measures do not show significant results.
El-Shagi and Kelly (2013)	Euro area	Eurostat, IMF, ECB, NCB	M3	2003–2013	Divisia monetary aggregates, Signals approach	Divisia aggregates are excellent predictors of debt crises, better capturing liquidity changes than simple sum measures.

Table A1. Cont.

Authors	Country	Data Sources	Components	Years	Research Design	Main Results
Fleissig et al. (2023)	Euro Area	ECB	M1, M2, M3	2000–2021	Fourier demand model, VAR models, Impulse response functions	Divisia monetary aggregates are robust to benchmark rates except during the global financial crisis, indicating differences from simple sum aggregates, particularly in elasticity and substitution responses.
Ghosh and Bhadury (2018)	India, Israel, Poland, the UK, the US	Bank of England, Bank of Israel, National Bank of Poland, Ramachandran et al. (2010), OECD database, Center for Financial Stability	M1, M2, M3, M4	1994–2017	Unit root test, Var parameter stability test, Bootstrapped Granger causality	Strong causality from Divisia money to exchange rates, particularly when compared to alternative monetary indicators like simple sum aggregates or interest rates.
Gogas et al. (2019)	Euro Area	CEPR, Bruegel	M1, M2, M3	2001–2018	Support Vector Machines	Divisia monetary aggregates significantly improve GDP forecasting accuracy compared to simple sum aggregates, demonstrating robust predictive power in various economic conditions and aggregation levels.
Hendrickson (2014)	US	Center for Financial Stability, St. Louis Federal Reserve FRED	M1, M2	1967–2012	Cointegrated VAR model, Granger causality tests, Backward-looking IS equation, Dynamic New Keynesian model	Divisia monetary aggregates provide better explanatory power for money demand stability, nominal income, price level, and the output gap compared to simple sum aggregates.
Keating et al. (2014)	US	FRED, BEA, CRB	M4	1960–2013	Structure VAR model, Impulse response functions, Granger causality tests	Divisia M4, as the policy indicator variable, solves the price puzzle without needing to add extraneous information to the model and remains informative even when interest rates are near zero.
Kelly et al. (2024)	Switzerland	Swiss National Bank data via DataStream	M3	1984–2019	MRN model, Bayesian VAR benchmark	Incorporating both conventional and Divisia monetary aggregates along with a short-term interest rate significantly enhances inflation forecasting accuracy over 12-, 24-, and 36-month horizons.
Leong et al. (2019)	Malaysia	Not mentioned, but possibly from Bank Negara Malaysia	M2	1991–2018	NARDL approach, ADF test, CUSUM, and CUSUM Squared tests	Divisia monetary aggregate and exchange rate changes significantly affect money demand, with asymmetric effects observed for currency appreciation in the long run. The model using Divisia money is stable and performs better in estimating the money demand function.

Table A1. Cont.

Authors	Country	Data Sources	Components	Years	Research Design	Main Results
Polat (2018)	Turkey	Central Bank of Turkey	M1, M2	2005–2016	Divisia monetary aggregates, Wavelet analysis, Structural VAR analysis	Divisia aggregates show high comovement with GDP at low frequencies and some predictive power for output and prices, but their theoretical advantages are not strongly supported by empirical evidence.
Puah and Hiew (2011)	Indonesia	IFS	M1, M2	1981–2005	VECM, Granger causality, residual, and cointegration test	Divisia M1 outperforms other aggregates in estimating stable money demand, exhibiting the fastest adjustment to long-run equilibrium.
Ramachandran et al. (2010)	India	Reserve Bank of India	M2, M3, L1	1993–2008	Divisia monetary aggregates, VECM, Cointegration tests	Divisia monetary aggregates outperformed simple sum measures in predicting inflation in India, suggesting they could be more valuable indicators for monetary policy during the post-liberalization period.
Restrepo-Tobón (2015)	UK	ONS, CIA’s World Fact Book, FRED, Bank of England, U.K. unit trusts	Not applicable	1965–2011	Habit-based asset pricing model, CC and DCC model	Risk adjustment is unnecessary when constructing Divisia monetary aggregates.
Sarwar et al. (2010)	Pakistan	State Bank of Pakistan (SBP), Statistical Bulletins, IMF	M0, M1, M2	1972–2007	Divisia monetary aggregates, ARDL approach, Cointegration, Error correction mechanism	Divisia monetary aggregate provides more realistic and stable money demand estimates, suggesting the SBP should switch from Simple sum to Divisia aggregates to enhance the effectiveness of monetary policy formulation.
Sengupta et al. (2024)	India	Reserve Bank of India, Handbook of Statistics on the Indian Economy	M1, M2, M3, M4	2001–2020	Divisia monetary aggregates, Granger causality tests, Cyclical correlations analysis	Divisia monetary aggregates Granger cause economic activity, but the link broke during the COVID-19 pandemic, highlighting the need for adaptive monetary policy strategies in response to unprecedented economic shocks.
Serletis and Gogas (2014)	US	AMFM	M1, M2	1967–2011	Johansen (1988) maximum likelihood approach, VAR model	Divisia aggregates better validate the long-run money demand function compared to simple-sum aggregates, offering more accurate and stable estimates.

Notes: The compilation includes only English-language papers after 2010.

Appendix B

Table A2. Monetary Asset Components (Hong Kong).

#	Asset	M1	M2	M3	Rate of Return
1	Currency	X	X	X	0%
2	Customers' Demand Deposits	X	X	X	0%
3	Customers' Savings Deposits ¹		X	X	Savings deposit rate (Banks)
4	Customers' Time Deposits ¹		X	X	6-month Time deposit rate (Banks)
5	Negotiable CDs (NCDs) ¹		X	X	364-day Exchange Fund Bills rate
6	Customers' Savings Deposits ²			X	Savings deposit rate (Banks)
7	Customers' Time Deposits ²			X	3-month Time deposit rate (Banks)
8	Negotiable CDs (NCDs) ²			X	364-day Exchange Fund Bills rate

Source: Hong Kong Monetary Authority (HKMA). ¹ Deposits that are issued with licensed banks and held by the public. ² Deposits are issued with restricted license banks and deposit-taking companies and held by the public.

Table A3. Monetary Asset Components (South Korea), where M3 is called Lf and M4 is called L.

#	Asset	M1	M2	Lf	L	Rate of Return
1	Currency	X	X	X	X	0%
2	Demand Deposits	X	X	X	X	0.3–0.6%
3	Transferable Saving Deposits	X	X	X	X	0.2–1.5%
4	Money Market Funds (MMFs)		X	X	X	MMFs (7-day)
5	Short-term Time and Savings Deposits		X	X	X	CDs (3-month)
6	Beneficiary Certificates		X	X	X	Yields of Treasury Bonds (1-year)
7	Marketable Financial Instruments ¹		X	X	X	Marketable Financial instruments rate
8	Short-term Financial Debentures		X	X	X	Short-term Financial debenture rate
9	Short-term Money in Trust		X	X	X	CDs (3-month)
10	Other Short-term Financial Assets ²		X	X	X	CDs (3-month)
11	Long-term financial instruments			X	X	Yields of Treasury Bonds (3-year)
12	Life Insurance Reserves			X	X	Yields of Treasury Bonds (3-year)
13	Liquid financial instruments ³				X	Yields of Treasury Bonds (1-year)

Source: Economic Statistics System (ECOS) from Bank of Korea. ¹ Marketable financial instruments include CDs (Certificates of Deposit), RPs (Repurchase Agreements), and cover bills. ² Other short-term financial assets include CMAs (Cash Management Accounts), foreign currency deposits with maturities of less than two years, bills issued by merchant banking corporations, and securities investment savings at investment trust companies. ³ Liquid financial instruments include government bonds & municipal bonds, financial instruments of other financial corporations, and corporate bonds & commercial papers.

Table A4. Monetary Asset Components (Singapore).

#	Asset	M1	M2	M3	Rate of Return
1	Currency	X	X	X	0%
2	Demand Deposits	X	X	X	0%
3	Fixed Deposits (Banks) ¹		X	X	3-month Fixed Deposits (Banks)
4	Negotiable CDs (NCDs, Banks)		X	X	3-month T-bill yield
5	Saving & Other Deposits (Banks) ²		X	X	Savings Deposits (Banks)
6	Fixed Deposits (Finance)			X	3-month Fixed Deposits (Finance)
7	Saving & Other Deposits (Finance)			X	Savings Deposits (Finance)

Source: Monetary Authority of Singapore (MAS). ¹ Including the POSB's fixed deposits; ² Including the POSB's savings deposits.

Table A5. Monetary Asset Components (Taiwan).

#	Asset	M1A	M1B	M3	Rate of Return
1	Currency	X	X	X	0%
2	Passbook Deposits	X	X	X	Passbook deposit rate *
3	Passbook Savings Deposits		X	X	Passbook savings deposit rate *
4	Time and Savings Deposits			X	2-yr Time deposit rate *
5	Foreign Currency Deposits			X	Foreign currency deposit rate (for US\$) *
6	Postal Savings Deposits			X	Passbook savings deposit rate of Chunghwa Post Co.
7	Negotiable CDs (NCDs)			X	91–180 days’ NCD rate in the secondary market
8	Repurchase Agreements			X	91–180 days’ Repos rate on Government bonds in the secondary market
9	MMFs and Government Deposits			X	Overnight Interbank Loan rate

Source: Central Bank of the Republic of China (Taiwan). * Rates offered by the Bank of Taiwan.

Table A6. ADF unit root test (Hong Kong).

<i>H</i> ₀ : Variable Has a Unit Root		
Variables	ADF Test Statistics	
	Levels	First Difference
NGDP	−0.8980029 (0.7896)	−21.45086 (0.00)
M1	−1.778881 (0.3915)	−4.618605 (0.00)
M2	−0.6147134 (0.8651)	−5.56583 (0.00)
M3	−0.6203887 (0.8638)	−5.545811 (0.00)
DM1	−1.778881 (0.3915)	−4.618605 (0.00)
DM2	−0.4127978 (0.9048)	−5.216365 (0.00)
DM3	−0.4176722 (0.9039)	−5.211939 (0.00)
R	−2.9608 (0.0387)	−8.729453 (0.00)

Note: The numbers in parentheses are *p*-values.

Table A7. Test Statistics from the Johansen Cointegration Test (Hong Kong).

Johansen-Procedure Cointegration Test:		
Test Type: Maximal Eigenvalue Statistic (Lambda Max), without Linear Trend and Constant in Cointegration.		
Variables	Hypothesis	Test Statistic
NGDP & M1	$r = 0$	23.02 *
	$r \leq 1$	4.55
NGDP & M2	$r = 0$	14.05
	$r \leq 1$	0.51
NGDP & M3	$r = 0$	14.12
	$r \leq 1$	0.50
NGDP & DM1	$r = 0$	23.02 *
	$r \leq 1$	4.55
NGDP & DM2	$r = 0$	15.29
	$r \leq 1$	0.14
NGDP & DM3	$r = 0$	15.43
	$r \leq 1$	0.13
NGDP & R	$r = 0$	9.09
	$r \leq 1$	0.94
Critical values	$r = 0$	17.95
	$r \leq 1$	8.18

* Indicate significance at 5%.

Table A8. ADF unit root test (South Korea).

<i>H</i> ₀ : Variable Has a Unit Root		
Variables	ADF Test Statistics	
	Levels	First Difference
NGDP	−1.691144 (0.4358)	−5.107981 (0.00)
M1	−0.5400484 (0.8810)	−2.823301 (0.0550)
M2	0.3280406 (0.9798)	−3.074774 (0.0285)
Lf	−0.8881352 (0.7926)	−3.926544 (0.0019)
L	−1.774415 (0.3937)	−3.747036 (0.0035)
DM1	−0.5450154 (0.8800)	−2.824143 (0.0549)
DM2	−0.2726098 (0.9266)	−2.918711 (0.0432)
DLf	−1.242655 (0.6581)	−3.49344 (0.0082)
DL	−1.907839 (0.3290)	−3.387125 (0.0114)
R	−2.060423 (0.2611)	−3.161076 (0.0224)

Note: The numbers in parentheses are *p*-values.

Table A9. Test Statistics from the Johansen Cointegration Test (South Korea).

Johansen-Procedure Cointegration Test:		
Test Type: Maximal Eigenvalue Statistic (Lambda Max), without Linear Trend and Constant in Cointegration.		
Variables	Hypothesis	Test Statistic
NGDP & M2	$r = 0$	5.94
	$r \leq 1$	2.13
NGDP & Lf	$r = 0$	11.76
	$r \leq 1$	0.72
NGDP & L	$r = 0$	13.47
	$r \leq 1$	02.34
NGDP & DM2	$r = 0$	6.61
	$r \leq 1$	2.54
NGDP & DLf	$r = 0$	12.04
	$r \leq 1$	1.02
NGDP & DL	$r = 0$	13.60
	$r \leq 1$	2.32
NGDP & R	$r = 0$	9.12
	$r \leq 1$	3.43
Critical values	$r = 0$	17.95
	$r \leq 1$	8.18

Table A10. ADF unit root test (Singapore).

<i>H</i> ₀ : Variable Has a Unit Root		
Variables	ADF Test Statistics	
	Levels	First Difference
NGDP	−2.145387 (0.2269)	−1.81814 (0.3720)
M1	−1.792058 (0.3849)	−3.230571 (0.0183)
M2	−3.272511 (0.0162)	−3.440986 (0.0097)
M3	−3.297754 (0.0150)	−3.387432 (0.0114)
DM1	−1.792058 (0.3849)	−3.230571 (0.0183)
DM2	−3.268824 (0.0164)	−3.425891 (0.0101)
DM3	−3.292385 (0.0153)	−3.375615 (0.0119)
R	−1.956108 (0.3067)	−3.661892 (0.0047)

Note: The numbers in parentheses are *p*-values.

Table A11. ADF unit root test (Taiwan).

<i>H</i> ₀ : Variable Has a Unit Root		
Variables	ADF Test Statistics	
	Levels	First Difference
NGDP	0.5175334 (0.9874)	−5.939329 (0.0000)
M1A	0.5485959 (0.9883)	−4.15461 (0.0008)
M1B	0.1234442 (0.9675)	−3.899677 (0.0020)
M2	1.475913 (0.9993)	−4.313209 (0.0004)
DM1A	1.463601 (0.9993)	−4.734366 (0.0001)
DM1B	0.7366827 (0.9930)	−3.908257 (0.0020)
DM2	0.7125017 (0.9925)	−4.063133 (0.0011)
R	−2.462664 (0.1248)	−4.720949 (0.0001)

Note: The numbers in parentheses are *p*-values.

Table A12. Test Statistics from the Johansen Cointegration Test (Taiwan).

Johansen-Procedure Cointegration Test:		
Test Type: Maximal Eigenvalue Statistic (Lambda Max), without Linear Trend and Constant in Cointegration.		
Variables	Hypothesis	Test Statistic
NGDP & M1A	<i>r</i> = 0	9.64
	<i>r</i> ≤ 1	0.07
NGDP & M1B	<i>r</i> = 0	11.66
	<i>r</i> ≤ 1	0.92
NGDP & M2	<i>r</i> = 0	12.6
	<i>r</i> ≤ 1	3.4
NGDP & DM1A	<i>r</i> = 0	19.92 *
	<i>r</i> ≤ 1	1.79
NGDP & DM1B	<i>r</i> = 0	12.34
	<i>r</i> ≤ 1	0.00
NGDP & DM2	<i>r</i> = 0	7.24
	<i>r</i> ≤ 1	0.53
NGDP & R	<i>r</i> = 0	7.15
	<i>r</i> ≤ 1	1.32
Critical values	<i>r</i> = 0	17.95
	<i>r</i> ≤ 1	8.18

* Indicate significance at 5%.

Table A13. Dynamic Factor Analysis and Residual Diagnostics (Hong Kong).

Factor Loadings		
	PC1	PC2
GDP	0.7071068	−0.7071068
M1	0.7071068	0.7071068
Correlations between VECM residuals and factor		
	Factor 1	Factor 2
Residuals of GDP	0.01043529	−0.2062272
Residuals of DM1	−0.17128507	0.1769230
Residual Diagnostics		
Multivariate Shapiro-Wilk test for normality:	<i>p</i> -value = 0.355	
Portmanteau Test (asymptotic)	<i>p</i> -value = 0.1054	
Multivariate ARCH-LM test:	<i>p</i> -value = 0.217	

Note: The results are the same for DM1 as due to the interest rates being zero for the components in M1, the simple sum and Divisia measures for narrow money in Hong Kong are the same.

Table A14. Dynamic Factor Analysis and Residual Diagnostics (Taiwan).

Factor Loadings		
	PC1	PC2
GDP	0.7071068	0.7071068
DM1A	0.7071068	−0.7071068
Correlations between VECM residuals and factor		
	Factor 1	Factor 2
Residuals of GDP	0.02935621	0.6809136
Residuals of DM1	0.15298498	−0.1841777
Residual Diagnostics		
Multivariate Shapiro-Wilk test for normality:	<i>p</i> -value = 0.08831	
Portmanteau Test (asymptotic)	<i>p</i> -value = 0.3667	
Multivariate ARCH-LM test:	<i>p</i> -value = 0.2461	

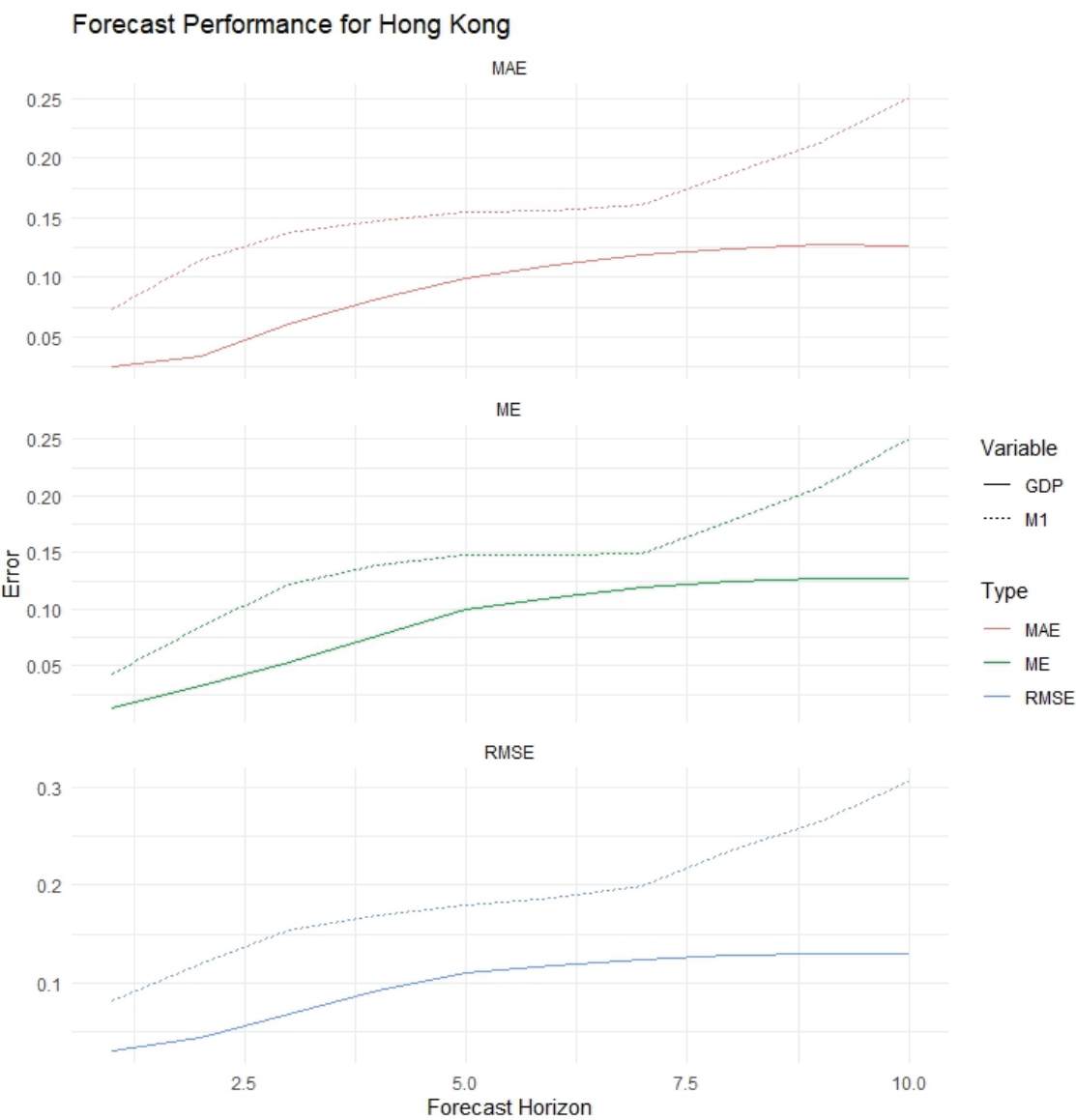


Figure A1. Forecast Performance Metrics for Hong Kong.

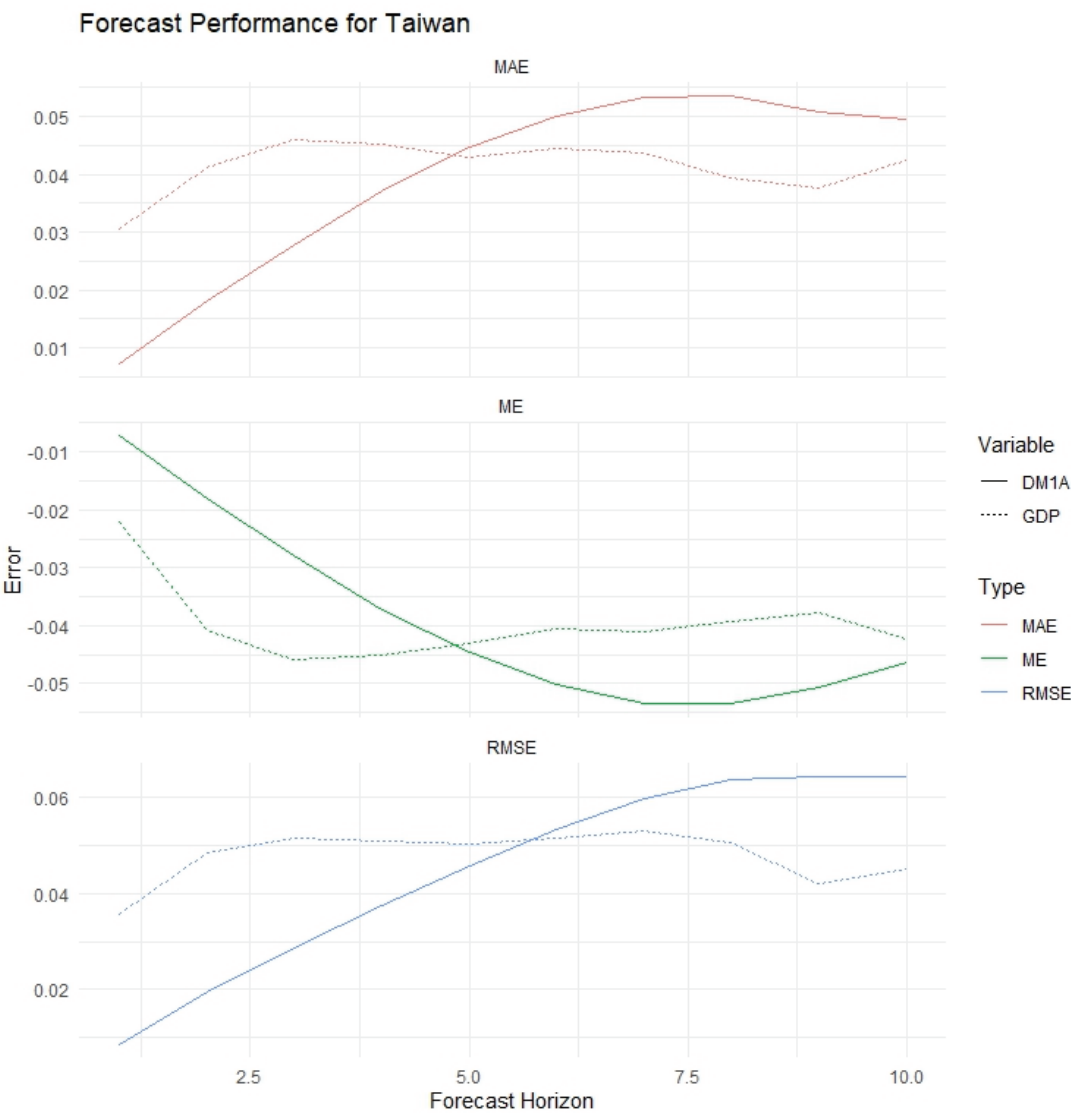


Figure A2. Forecast Performance Metrics for Taiwan.

- Notes
- ¹ The Bank of Korea's issue note focused on studying M2, precisely matching components with corresponding interest rates. The matching was done as follows: currency (0%), demand deposits (demand deposit rate), short-term time deposits with maturities (pure savings deposits), short-term savings deposits (savings deposit rate), certificates of deposit (certificate of deposit rate), repurchase agreements (repurchase agreement rate), and financial debentures (financial debenture rate). Other items, particularly those not held for transactional purposes, were likely excluded from the analysis, indicating a prioritization of components with clearly defined interest rates. We thank Dr. Park for his kind response to our email inquiry.
- ² We would also like to extend our gratitude to the former Financial Supervisory Service employee who provided helpful insights into the BOK's data collection process.
- ³ Lee (2016) analyzed the 3-year government bond yield as a variable that increased the risk-weighted asset ratio in life insurance companies following the global financial crisis of 2008. This suggests that the 3-year government bond yield influences the interest rate decisions of life insurance companies and can be used as a proxy interest rate.

References

- Akaike, Htrotugu. 1973. Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika* 60: 255–65. [CrossRef]
- Alkhareif, Ryadh, and Moayad Al-Rasasi. 2021. Estimating the money demand function for Saudi Arabia using Divisia monetary aggregate. *Applied Economics* 53: 4823–34. [CrossRef]
- Anderson, Richard G., and Barry E. Jones. 2011. A comprehensive revision of the U.S. monetary services (Divisia) indexes. *Federal Reserve Bank of St. Louis Review* 93: 325–60. [CrossRef]
- Barnett, William A. 1978. The user cost of money. *Economic Letters* 1: 145–49. [CrossRef]
- Barnett, William A. 1980. Economics monetary aggregates: An application of index number and aggregation theory. *Journal of Econometrics* 14: 11–48. [CrossRef]
- Barnett, William A. 1981. The new monetary aggregates: A comment. *Journal of Money, Credit, and Banking* 13: 485–89. [CrossRef]
- Barnett, William A. 1982. The optimal level of monetary aggregation. *Journal of Money, Credit, and Banking* 14: 687–710. [CrossRef]
- Barnett, William A. 2010. Monetary aggregation. In *Monetary Economics*. Edited by Steven N. Durlauf and Lawrence E. Blume. The New Palgrave Economics Collection. London: Palgrave Macmillan.
- Barnett, William A. 2016. Friedman and Divisia monetary measures. In *Milton Friedman: Contributions to Economics and Public Policy*. Edited by Robert A. Cord and J. Daniel Hammond. Oxford: Oxford University Press. [CrossRef]
- Barnett, William A., and Neepe Gaekwad. 2021. *Multilateral Divisia Monetary Aggregates for the Euro Area*. Working Papers Series in Theoretical and Applied Economics 202108. Lawrence: Department of Economics, University of Kansas.
- Barnett, William A., and Van H. Nguyen. 2021. Constructing Divisia monetary aggregates for Singapore. *Journal of Risk and Financial Management* 14: 370. [CrossRef]
- Barnett, William A., Douglas Fisher, and Apostolos Serletis. 1992. Consumer theory and the demand for money. *Journal of Economic Literature* 30: 2086–119. Available online: <https://www.jstor.org/stable/2727974> (accessed on 11 February 2023).
- Barnett, William A., Giovanni Bella, Taniya Ghosh, Paolo Mattana, and Beatrice Venturi. 2022a. Is policy causing chaos in the United Kingdom? *Economic Modelling* 108: 105767. [CrossRef]
- Barnett, William A., Jia Liu, Ryan S. Mattson, and van den Jeff Noort. 2013. The new CFS Divisia monetary aggregates: Design, construction, and data sources. *Open Economies Review* 24: 101–24. [CrossRef]
- Barnett, William A., Kun He, and Jingtong He. 2022b. Consumption loan augmented Divisia monetary index and China monetary aggregation. *Journal of Risk and Financial Management* 15: 447. [CrossRef]
- Barnett, William A., Soumya S. Bhadury, and Taniya Ghosh. 2016. A SVAR approach to evaluation of monetary policy in India: Solution to the exchange rate puzzles in an open economy. *Open Economies Review* 27: 871–93. [CrossRef]
- Belongia, Michael T. 1996. Measurement matters: Recent results from monetary economics reexamined. *Journal of Political Economy* 104: 1065–83. [CrossRef]
- Belongia, Michael T., and Peter N. Ireland. 2016. Money and output: Friedman and Schwartz revisited. *Journal of Money, Credit and Banking* 48: 1223–66. [CrossRef]
- Belongia, Michael T., and Peter N. Ireland. 2021. Targeting Nominal Income under the Zero Lower Bound: The Case of the Bank of England. Available online: <https://centerforfinancialstability.org/amfm/studies/ukngdp2021.pdf> (accessed on 8 August 2024).
- Belongia, Michael T., and Peter N. Ireland. 2022. Strengthening the second pillar: A greater role for money in the ECB's strategy. *Applied Economics* 54: 99–114. [CrossRef]
- Benchimol, Jonathan. 2016. Money and monetary policy in Israel during the last decade. *Journal of Policy Modeling* 38: 103–24. [CrossRef]
- Berar, R., and Jenny Owladi. 2013. Amendments to Divisia money series. *Monetary and Financial Statistics*. January 1–2. Available online: <https://escocoe-website.s3.amazonaws.com/wp-content/uploads/2019/01/30141731/Amendments-to-Divisia-money-series-by-Rajveer-Berar-and-Jenny-Owladi.pdf> (accessed on 8 August 2024).
- Bhatnagar, Aryaman. 2022. Divisia Aggregates and Monetary Policy in India. *unpublished paper*.
- Binner, Jane M., Alicia M. Gazely, and Shu H. Chen. 2002. Financial innovation and Divisia monetary indices in Taiwan: A neural network approach. *The European Journal of Finance* 8: 238–47. [CrossRef]
- Binner, Jane M., and Logan J. Kelly. 2017. Modelling money shocks in a small open economy: The case of Taiwan. *The Manchester School* 85: 104–20. [CrossRef]
- Bissoondeed, Rakesh K., Barry E. Jones, Jane M. Binner, and Andrew W. Mullineux. 2010. Household-sector money demand for the UK. *The Manchester School* 78: 90–113. [CrossRef]
- Boughton, James H. 1981. Money and its substitutes. *Journal of Monetary Economics* 8: 375–86. [CrossRef]
- Brill, Maximilian C., Dieter Nautz, and Lea Sieckmann. 2021. Divisia monetary aggregates for a heterogeneous euro area. *Empirica* 48: 247–78. [CrossRef]
- Chen, Wenjuan, and Dieter Nautz. 2015. The Information Content of Monetary Statistics for the Great Recession: Evidence from Germany (No. 2015-027). SFB 649 Discussion Paper. Available online: <https://www.econstor.eu/bitstream/10419/119434/1/826639127.pdf> (accessed on 8 August 2024).
- Chetty, V. Karuppan. 1969. On measuring the nearness of near-moneys. *American Economic Review* 59: 270–81. Available online: <http://www.jstor.org/stable/1808957> (accessed on 11 February 2023).

- Chrystal, Alec K., and Ronald MacDonald. 1994. Empirical evidence on the recent behavior and usefulness of simple-sum and weighted measures of the money stock. *Federal Reserve Bank of St. Louis Review* 76: 73–109. [CrossRef]
- Darvas, Zsolt. 2015. Does money matter in the euro area? Evidence from a new Divisia index. *Economics Letters* 133: 123–26. [CrossRef]
- Diewert, Erwin W. 1976. Exact and superlative index numbers. *Journal of Econometrics* 4: 115–45. [CrossRef]
- El-Shagi, Makram, and Logan J. Kelly. 2013. Liquidity in the Liquidity Crisis: Evidence from Divisia Monetary Aggregates in Germany and the European Crisis Countries. *Economics Bulletin* 34: 63–72. Available online: <http://www.accessecon.com/Pubs/EB/2014/Volume34/EB-14-V34-I1-P7.pdf> (accessed on 6 August 2024).
- Engle, Robert F., and Clive W. J. Granger. 1987. Co-integration and error correction: Representation, estimation, and testing. *Econometrica* 55: 251–76. [CrossRef]
- Feige, Edgar, and Douglas Pearce. 1977. The substitutability of money and near-monies: A survey of the time-series evidence. *Journal of Economic Literature* 15: 439–69. Available online: <https://www.jstor.org/stable/2723210> (accessed on 11 February 2023).
- Fisher, Irving. 1922. *The Making of Index Numbers: A Study of Their Varieties, Tests, and Reliability*. Boston and New York: Houghton Mifflin Company.
- Fleissig, Adrian R., Barry E. Jones, and Zsolt Darvas. 2023. Euro area monetary asset demand and Divisia aggregates. *The European Journal of Finance* 29: 1885–912. [CrossRef]
- Friedman, Milton, and Anna J. Schwartz. 1963. *A Monetary History of the United States, 1867–1960*. Princeton: Princeton University Press. Available online: <http://www.jstor.org/stable/j.ctt7s1vp> (accessed on 23 September 2024).
- Ghosh, Taniya, and Soumya Bhadury. 2018. Money’s causal role in exchange rates: Do Divisia monetary aggregates explain more? *International Review of Economics & Finance* 57: 402–17. [CrossRef]
- Gogas, Periklis, Theophilos Papadimitriou, and Emmanouil Sofianos. 2019. Money neutrality, monetary aggregates and machine learning. *Algorithms* 12: 137. [CrossRef]
- Habibullah, Muzafar S. 1999. *Divisia Monetary Aggregates and Economic Activities in Asian Developing Economies*, 1st ed. London: Routledge. [CrossRef]
- Hahm, Jeong Ho, and Jun Tae Kim. 2000. The signals from Divisia money in a rapidly growing economy. In *Divisia Monetary Aggregates*. Edited by Michael T. Belongia and Jane M. Binner. London: Palgrave Macmillan. [CrossRef]
- Hendrickson, Joshua R. 2014. Redundancy or mismeasurement? A reappraisal of money. *Macroeconomic Dynamics* 18: 1437–65. [CrossRef]
- International Monetary Fund. 2000. *Monetary and Financial Statistics Manual*. Washington, DC: International Monetary Fund.
- Ishida, Kazuhiko. 1984. Divisia monetary aggregates and the demand for money: A Japanese case. *Bank of Japan Monetary and Economic Studies* 2: 49–80. Available online: <https://www.imes.boj.or.jp/research/papers/english/me2-1-2.pdf> (accessed on 7 August 2024).
- Jha, Raghbendra, and Ibotombi S. Longjam. 2008. A Divisia type saving aggregate for India. *Macroeconomics and Finance in Emerging Market Economies* 1: 51–66. [CrossRef]
- Johansen, Søren. 1988. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12: 231–54. [CrossRef]
- Johansen, Søren, and Katarina Juselius. 1990. Maximum likelihood estimation and inference on cointegration—With applications to the demand for money. *Oxford Bulletin of Economics and Statistics* 52: 169–210. [CrossRef]
- Keating, John W., Logan J. Kelly, and Valcarcel J. Valcarcel. 2014. Solving the price puzzle with an alternative indicator of monetary policy. *Economics Letters* 124: 188–94. [CrossRef]
- Kelly, Logan J., Jane M. Binner, and Jonathan Tepper. 2024. Do monetary aggregates improve inflation forecasting in Switzerland? *Journal of Management Policy and Practice* 25: 124–33. [CrossRef]
- Korea Institute of Finance. 2016. Discussion on the utility of Divisia monetary aggregates. *Weekly Financial Brief* 25: 16–17. (In Korean).
- Kumah, Emmanuel O. 1989. Monetary Concept and Definitions. IMF Working Paper, 89/92. Available online: <https://ssrn.com/abstract=885054> (accessed on 7 June 2024).
- Lee, Seok Ho. 2016. A comparative analysis of managerial performance between financial capital-affiliated and industrial capital-affiliated insurance companies. In *KIF Financial Analysis Report 2016-02*. Seoul: Korea Institute of Finance. (In Korean)
- Leong, Choi-Meng, Chin-Hong Puah, Evan Lau, and Abu M. Shazali. 2019. Asymmetric effects of exchange rate changes on the demand for divisia money in Malaysia. *Journal of International Studies* 12: 52–62. [CrossRef]
- Michelson, Noam, and Tanya Suhoy. 2014. Weekly monitoring of financial systems in transition: A financial conditions index and a credit conditions index—the Israeli case. Paper presented at World Finance Conference, Venice, Italy, July 2–4.
- Moroney, John R., and Barry J. Wilbratte. 1976. Money and money substitutes: A time series analysis of household portfolios. *Journal of Money, Credit and Banking* 8: 181–98. [CrossRef]
- Park, Kyung Hoon, and Yeon Jeong Shim. 2018. Analysis of the relationship between money supply and economic conditions. In *Issue Note 2018-1*. Seoul: Bank of Korea. (In Korean)
- Polat, Umurcan. 2018. Divisia and simple sum monetary aggregates: Any empirical relevance for Turkey? *Journal of Central Banking Theory and Practice* 7: 175–206. [CrossRef]
- Puah, Chin-Hong, and Lee-Chea Hiew. 2011. Financial liberalization weighted monetary aggregates and money demand in Indonesia. *Labuan Bulletin of International Business and Finance (LBIBF)* 8: 76–93. [CrossRef]
- Ramachandran, M., Rajib Das, and Binod B. Bhoi. 2010. The Divisia monetary indices as leading indicators of inflation. *Indian Economic Journal* 58: 3–30. [CrossRef]

- Restrepo-Tobón, Diego A. 2015. Evidence that risk adjustment is unnecessary in estimates of the user cost of money. *Ecos de Economía* 19: 49–70. [CrossRef]
- Sarwar, Haroon, Zakir Hussain, and M. Sarwar Awan. 2010. Money demand function for Pakistan (Divisia approach). *Pakistan Economic and Social Review* 48: 1–20. Available online: <http://www.jstor.org/stable/41762411> (accessed on 8 August 2024).
- Sengupta, Anirban, Apostolos Serletis, and Libo Xu. 2024. Divisia monetary aggregates for India. *Studies in Nonlinear Dynamics & Econometrics*. [CrossRef]
- Serletis, Apostolos, and Periklis Gogas. 2014. Divisia monetary aggregates, the great ratios, and classical money demand functions. *Journal of Money, Credit and Banking* 46: 229–41. Available online: <http://www.jstor.org/stable/42920078> (accessed on 7 August 2024). [CrossRef]
- Shih, Yen C. 2000. Divisia monetary aggregates for Taiwan. In *Divisia Monetary Aggregates*. Edited by Michael T. Belongia and Jane M. Binner. London: Palgrave Macmillan. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

The NGOs Efficiency: A PROMETHEE Approach

Susana Álvarez-Otero * and Emma Álvarez-Valle

Department of Business Administration, Faculty of Economics and Business, University of Oviedo,
33006 Oviedo, Spain; emmaav@hotmail.es

* Correspondence: saotero@uniovi.es; Tel.: +34-985182209

Abstract: The current economic and political crisis has brought about a change in the environment in which non-governmental development organisations (NGDOs) have traditionally operated. This change can be summed up as a reduction in the funds they receive and an increase in the population they must serve. The need then arises to have mechanisms that allow an analysis of the good work performed by the NGDOs. Knowing the efficiency of the NGDOs in the management of their previous projects can contribute towards improving their future achievements. The aim of this research is to establish some objective indicators that allow an evaluation of the efficiency of these organisations. Firstly, a detailed analysis of the regulation of the three agencies is conducted (Spanish-AECID, European-EuropeAid, and American-USAID). This allows us to synthesise the indicators of good performance of the NGDO based on the study of the eligibility criteria of public donors. The research concludes with the study of the efficiency following the Promethee Approach. Our results reveal that 44.6% of the NGDOs (33 out of the 74 studied) operate inefficiently, compared to 29.7%, which are efficient.

Keywords: non-governmental development organisation (NGDO); performance; indicators; cooperation; donor; promethee

1. Introduction

In recent years, the economic crises and the publication in the media of dubious practices by Non-Governmental Development Organisations (NGDOs) have put organisations dedicated to development cooperation in trouble. It is necessary to strengthen these non-profit organisations with donors and with society in general. In addition, provide their managers with tools that allow them to identify the areas in which they must act to improve their results, making the most of their resources. Efficiency studies cover both needs: on the one hand, they make it possible to identify which entities are the best performers, legitimising them in relation to society, and, on the other hand, they give organisations a vision for their operation and possible areas for improvement.

There are several studies dedicated to efficiency in NGDOs. However, the limitations that third-sector entities present, mainly due to the absence of a benefit that serves as an indicator and the use of peculiar resources, such as free donor financing, public and private, or the use of volunteers, result in not having objective and agreed items to measure efficiency. There is not even a doctrinal consensus on the very notion of efficiency in the field of development cooperation, nor on what the outputs are or, the objectives that these entities pursue.

This paper aims to study the efficiency in the development cooperation sector, define the objectives and resources of the NGDOs, and find general indicators that allow verifying to what extent the entities are achieving their goals and using their resources to be able to apply them to an efficiency study.

The efficiency indicators are inferred from those used by public donors when they select the NGDOs to receive funding, the so-called eligibility criteria published in different norms and laws. Thus, efficiency is measured in this paper using the indicators collected in

Citation: Álvarez-Otero, Susana, and Emma Álvarez-Valle. 2024. The NGOs Efficiency: A PROMETHEE Approach. *Journal of Risk and Financial Management* 17: 382. <https://doi.org/10.3390/jrfm17090382>

Academic Editor: Thanasis Stengos

Received: 18 July 2024

Revised: 15 August 2024

Accepted: 21 August 2024

Published: 26 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

the eligibility criteria of three of the main cooperation agencies, the Spanish Agency for International Development Cooperation (AECID), the European Agency (EuropeAid), and the Agency for the United States for International Development (USAID).

It is a novel method of fixing identifiers because, up to now, in the studies analysed, they were established either by the authors themselves or by means of surveys conducted with NGDO personnel or agency technicians. Efficiency indicators are inferred from those used by public donors when selecting NGDOs to receive funding, the so-called eligibility criteria published in different regulations and laws in force. This concept is more appropriate because it comes precisely from the perspective of the donor and therefore determines its funding and possibilities for future projects.

A very important novelty of this research, which fills a gap in the literature on this subject, is that we measure efficiency using the indicators included in the eligibility criteria of three of the main cooperation agencies mentioned before. This is a novel method of establishing identifiers, as until now, in the studies analysed, they were established either by the authors themselves or by means of surveys of NGDO staff or agency technicians. The fact that the indicators are obtained from donor regulations allows them to be more general and objective than those traditionally used.

One of the first facts found in this study is the opacity surrounding the eligibility criteria. The different criteria, the indicators used to corroborate compliance, and the sources from which the donors extract the data to quantify the indicator are scattered across different standards without establishing a clear relationship between them. The weights are generally available for the criteria without disaggregating the weight of each indicator.

Given this opacity and in order to know if the indicators published by donors were sufficient to determine the award of a public grant to a NGDO, a simulation of granting public funding was applied. For this, the 74 NGDOs that in 2016 (up to 2018–2023) belonged to the Spanish Coordinator were taken. With the donor variables, a study of the efficiency of the coordinator's NGDOs is carried out, using a technique that is an adaptation of the Promethee multi-criteria method known as PROMETHEE Productivity Analysis (PPA) developed by Ishizaka et al. (2017) to study the efficiency of UK universities. Our efficiency study reveals that 44.6% of the NGDOs (33 out of the 74 studied) operate inefficiently, compared to 29.7%, which are efficient.

We dedicate the following section to the literature review. The third section is orientated towards inferring, from donor regulations, the efficiency indicators. For this, the eligibility criteria of three public agencies are analysed: the Spanish Cooperation Agency (AECID), the European Agency (EuropeAid), and the American Agency (USAID). For each agency, the criteria, indicators, verification sources, and associated weights are extracted and reported in its regulations. This study shows that the criteria related to the narrative of the candidate project to be subsidised are, apparently, the most valued by donors, but a pre-selection is deduced, the existence of an internal tide of criteria and indicators, which donors do not specify.

The fourth section begins by combining the indicators established by the three donors, analysed in the previous section, with the methodology of Promethee applied to the list of organisations that have received the most public grants. The paper ends with the conclusions of this research, taking into account the results obtained in the previous section, including limitations of the paper and future research.

2. Literature Review

The first attempts to develop performance indicators for non-profit organisations were based on the adaptation of those established for for-profit organisations. Despite doctrinal efforts to reach a consensus, as early as 1998, Forbes (1998) showed how difficult it was to find valid indicators, since measures such as profitability or return per shareholder, which made it possible to evaluate the performance of for-profit organisations, were not easy to translate to the non-profit world.

Measuring the performance of third sector entities is subject to a series of limitations already known in the academic literature. The absence of a benefit that serves as an indicator, the search for rewards with qualitative and intangible features, the performance of material and immaterial activities that are difficult to quantify, the existence of volunteer staff, or the difficulty of measuring the impact of their actions mean that neither what performance should be measured, nor the approach to measure it, nor the indicators to use for it, nor even what is the justification for measuring that performance (Cordery and Sinclair 2013; Moxham 2014; Bach-Mortensen and Montgomery 2018) is clarified. Adding to these difficulties is the fact that there is a lack of official data on the third sector, especially on NGOs, due to the limitation of having to quantify their performance in third countries. In addition, the strong dependence of NGOs on donations and the need to present a good image in order to continue to have the support of their funders, together with the lack of economic experts within these entities, can lead to misreporting, an incorrect imputation of accounting items that distort the operability of the data collected (Pérez Martín et al. 2012).

The limitations listed above mean that most of the published works approach the study of the efficiency of third sector entities from a common approach (Lecy et al. 2012; Cordery and Sinclair 2013; Mouchamps 2014), using various tools, all of them coming from the for-profit sphere, so that their validity for the third sector is questioned, and with indicators of achievement established subjectively by the authors of the study, generally based on the data available to carry it out. In this sense, Mouchamps (2014) states that traditional performance measurement tools (in his study he focused on three of them: the balanced scorecard, the social return on investment, and the Global Reporting Initiative) would need to be adapted in such a way to the peculiarities of the third sector that, if this could be performed successfully, we would be faced with a completely new tool.

Among the studies dedicated to analysing the performance of NGOs or the third sector in general (excluding, therefore, those dedicated exclusively to social action or humanitarian aid organisations), there are studies dedicated to analysing the efficiency of one of the activities carried out by the organisations or of one aspect of their activity. Thus, Ribar and Wilhelm (2002) analysed the efficiency in fundraising and in carrying out interventions by American NGOs, concluding that efficiency in one of the activities detracts from efficiency in the other. Hernangómez Barahona et al. (2006) studied the objective pursued by NGOs from the point of view of satisfying the interests that stakeholders place in the organisation and, based on this objective of creating value for participants, analysed the efficiency of Spanish NGOs. García Cebrián and Cerbós (2007) studied the efficiency of NGOs in attracting resources; the results obtained showed that the influence of size, measured by NGO income, is very relevant in their efficiency and that the NGOs that specialise in public funds or private funds obtain better results.

Berber et al. (2011) studied the efficiency of fundraising and service provision in non-profit organisations, taking as their objective the social benefit generated by these organisations. In their study, they state that to date no rigorous method has been presented to analyse the efficiency of non-profit organisations, with potential donors being satisfied with analysing the ratio of funds raised to administrative expenses, concluding that there are service provision costs that the organisations did not account for in order to improve their image in the eyes of potential donors, which influenced the measurement of efficiency.

Martín Pérez et al. (2015) use DEA to study the efficiency of Spanish NGOs' development interventions. They conclude that efficiency is to some extent independent of the specific environment where the actions are implemented and depends, to a greater extent, on the type of cooperation instrument through which the aid has been channelled. They also conclude that the country results highlight the importance of focussing on smaller areas of action.

Gálvez-Rodríguez et al. (2016) studied the efficiency of Colombian NGOs, focussing on transparency as a means to legitimise these entities. They conclude that only 16.83% of the organisations studied are efficient—those that employ the fewest workers and volunteers and those that receive the least public funding. They also note a strong correlation between

the transparency of the organisation, especially financial and performance transparency, and its efficiency, with the influence of governance transparency and the profile of the organisation on the organisation's efficiency not being significant.

Ferreira et al. (2017) focus on the study of fundraising activity and how the quality and quantity of fundraising can influence the performance of NGDOs. They conclude that the greater the diversity of funders, the lower the dependence of NGDOs on their donors, providing management autonomy that favours a better performance of the entity.

There are also studies dedicated to analysing the methodology used to study organisational efficiency and the most appropriate indicators for this purpose. Kaplan (2001) attempted to adapt the balanced scorecard of the for-profit sector to non-profit organisations. The same effort was made by González Quintana and Molina (2008) for Spanish organisations and Greiling (2010) for German third sector organisations. However, the use of the balanced scorecard has not taken off in the non-profit sector (Retolaza 2010), questioning whether the dimensions used to measure the performance of NGOs are appropriate, due to the fact that they do not include the evaluation of the decision-making process and due to the difficulty of evaluating the indicators established by the balanced scorecard to measure the outputs obtained and the necessary inputs for this (Mouchamps 2014). Another criticism is that it tempts managers to set unambitious targets in order to be able to publicise the success of their achievements (Bisbe and Barrubés 2012).

The tool commonly used by donors for the design and monitoring of interventions is the logical framework approach, which has also been widely criticised for its rigidity, to the extent that it has become a mere administrative procedure rather than a method of planning by objectives (Sainz 2007; Cordery and Sinclair 2013).

Mouchamps, in 2014, made an attempt to find the most suitable tool for measuring the performance of third sector organisations and came to the overwhelming conclusion that "kitchen scales (tools designed for the for-profit sector) are being used to weigh elephants (non-profit organisations)" and that specific systems need to be designed for the third sector. However, there is an absence of studies examining how such systems should be developed (Moxham 2014). Bach-Mortensen and Montgomery (2018) went as far as to state that non-profit entities do not analyse their performance due to a lack of consistent guidelines on how such an evaluation should be carried out, with what criteria and indicators.

As we have seen, there is no agreement on which tools, criteria, and indicators are the most appropriate for analysing the good work of third-sector entities and, therefore, of NGOs. However, it is recognised in the studies analysed (Medina Rey 2009), as well as by the cooperation agencies themselves (Comisión Europea 2016; USAID 2012; Resolution of 17 September 2013, of the Presidency of the Spanish Agency for International Development Cooperation), that the regulations and indicators established by donors to manage and choose the organisations to which they grant subsidies are rigorous and guarantee the good work of these organisations. Taking this statement into account, that is, if the criteria used by donors to select the NGDOs that will receive a grant are strict and guarantee the good work of the chosen organisations, they could serve as indicators of the good work of these entities and could give rise to indicators that are not only valid but also accepted for the third sector. Consequently, by analysing donor eligibility criteria and the indicators used by public donors to grant subsidies to some entities or others, it will be possible to obtain performance indicators for NGDOs that will allow us to classify entities according to their better or worse performance. We will devote the next section of this paper to this analysis.

3. Analysis of the Three Agencies

In this section, we establish a series of NGDO performance indicators based on the study of the eligibility criteria of public donors. For this purpose, an analysis of the regulations of three cooperation agencies has been carried out: the Spanish Agency for International Development Cooperation (AECID), the European Agency (Europe-Aid), and the United States Agency for International Development (USAID). This study has allowed us to synthesise the indicators of good performance used by each one of the agencies. By

comparing the criteria, indicators, requirements, weightings, and sources of verification used by the three agencies, we will deduce which criteria and indicators can be used to verify the internal performance of NGDOs. These common indicators will allow us to demonstrate the performance of NGDOs to any public donor and, by extension, to the stakeholders of these entities.

3.1. Analysis of the AECID's Eligibility Criteria

From the analysis of the general regulatory framework to which AECID grants are subject, it can be concluded that the eligibility criteria are defined in an imprecise and unclear manner, scattered throughout the regulations, and without establishing precise relationships between the criteria, the indicators used to verify compliance, and the sources of verification used for each indicator.

In all the regulations, there is only one criterion relating to the entity's performance, which is to have a structure capable of sufficiently guaranteeing the fulfilment of its objectives. However, the indicators used by the donor to analyse compliance with this sufficiency criterion must be deduced from the sources of verification listed in article 7 (f) of Royal Decree 193/2015 (BOE 2015), regulating the prescriptive registration in the AECID Register, as the agency requires the organisation chart of the organisation and its delegations, the number of partners, the staff hired, and the number of volunteers and development workers dedicated to activities related to international development cooperation. In sections (g) and (h) of the same article, and without relating them directly to any criteria, a certificate is requested accrediting the funds received from membership fees, donations, and income generated by private fundraising activities that could finance part of the entity's activity, and the annual report corresponding to the last year of activity, if any.

All the indicators listed, although not explicitly stated in the regulation, are clearly related to the analysis of the material and human resources that the entity possesses and its capacity to carry out the development actions it undertakes and can therefore be considered indicators of the criterion of sufficient structure.

This lack of legislative clarity in establishing the indicators used to verify compliance with the criterion of having a structure capable of sufficiently guaranteeing the fulfilment of its objectives means that it is not possible to be sure exactly which are the real indicators used by the agency to verify this sufficiency. It is not known whether it takes these data in an absolute sense or whether it takes into account other indicators derived from them. Nor does the legislation make explicit what requirements these indicators must meet in order to consider that the entity has sufficient structure (a minimum number of partners, staff, projects carried out, etc., a reference threshold that sets the boundary between a bad and a good structure). It is known that the weight given by law to these requirements is 100%, as failure to meet the criterion of having a sufficient structure will result in refusal of registration in the AECID Register. Therefore, to calculate the weight given to each indicator of the criterion, this 100% is divided among all indicators (see Table 1).

In the general regulations, the rules governing subsidies according to the type of development cooperation action (agreement, project, or action) establish four eligibility criteria, three of which refer to the formulation of the specific intervention to be subsidised: the quality and effectiveness of the aid, the complementarity of the project with the objectives of Spanish cooperation and the quality of the formulation, viability and sustainability of the proposed actions, and only one criterion referring to the internal functioning of the NGDO, institutional solvency, which is weighted with 20% of the total of the four criteria.

Furthermore, it is pointed out that, in the case of bidding for project subsidies, bidding NGDOs must have carried out at least five cooperation projects in the last eight years, which is why a new criterion is identified: the entity's previous experience, with the indicator being the number of projects carried out with a formal weighting of 100%. In the case of tenders for agreements, the NGDO is required to be qualified, a qualification requirement to which we will refer later (article 5 Order AEC/2909/2011 (BOE 2011)).

Table 1. Indicators from AECID’s general regulations.

	Indicator	Stage	Weight	Total Weight
1	Organisation chart	AECID Registration	12.50%	100%
2	Number of members	AECID Registration	12.50%	
3	Number of workers hired	AECID Registration	12.50%	
4	Number of volunteers	AECID Registration	12.50%	
5	Number of development workers	AECID Registration	12.50%	
6	Volume of private funds received	AECID Registration	12.50%	
7	Volume of public donations received	AECID Registration	12.50%	
8	Annual report of activities	AECID Registration	12.50%	

Source: Own elaboration.

After analysing the general regulations, it remains to study the specific regulations contained in the calls for proposals for each of the agency’s funding instruments. The AECID annually calls for grants for the implementation of projects; every four years it calls for agreements; and periodically calls for development cooperation actions.

Taking into account all the indicators established by the agency in the different funding instruments analysed, for the institutional solvency criterion, the only one referring to good internal functioning, the following indicators can be established: 1. having the following published on the organisation’s website: activity reports, audited accounts, and the composition of the governing bodies; 2. the number of employees according to the entity’s funds; 3. indicators relating to sources of funding; 4. indicators referring to the quality of justification of previous AECID subsidies; 5. the existence of rules of conduct for employees; 6. the existence of mechanisms for processing complaints and suggestions; and 7. the alliances made.

In both calls for projects and agreements, the criterion of institutional solvency has a weight for the AECID of 20% in the evaluation (article 9 of Order AEC/2909/2011 (BOE 2011)), but it is not known what specific weight the agency gives to each of the indicators of this solvency, so for the purposes of this study, it is considered that they all have the same weight, i.e., each indicator listed counts 1/7 of the 20% (2.86%) corresponding to each solvency indicator.

In addition, some of the indicators listed involve data that is not publicly available, so the associated indicators 4, 5, 6, and 7 should not be taken into account. In summary, Table 2 shows the indicators extracted from the reading of Order AEC/2909/2011 (BOE 2011) as well as the different resolutions calling for grants for the three AECID cooperation instruments, except for the one referring to the required qualification.

Table 2. Indicators from Order AEC/2909/2011 (BOE 2011) and the resolutions on the call for subsidies.

	Indicator	Stage	Weight
9	Transparency	To have the activity report published on the website	0.95%
10		To have the annual accounts published on the website	0.95%
11		To have the details of the members of the governing bodies published on the website	0.95%
12	Number of workers according to their funding	Scale of subsidies	2.86%
13	Sources of funding	Scale of subsidies	2.86%
14	Partnerships made	Scale of subsidies	2.86%

Source: Own elaboration.

The procedure to obtain the qualification is divided into two phases. In the first phase, it is analysed whether the organisation verifies a series of general requirements; without the accreditation of these requirements, it will not pass to the next phase and qualification will be denied. The second phase involves subjecting the organisation to a quantitative and qualitative assessment of its performance, giving each item analysed a certain score.

In order not to be exhaustive in listing all the indicators included in the resolution regulating the rating of entities, Tables 3 and 4 list only those indicators included in the assessment of the rating that are related to the internal functioning of the entity, leaving out the formal indicators and those whose verification requires unpublished data. Nor are the minimum compliance thresholds established by the agency to consider the indicator passed taken into account, since, as they are established for large NGOs, if these thresholds were taken into account, they would cease to be valid as indicators of good performance for medium-sized and small NGOs.

Table 3. Indicators collected in the first qualification phase.

Indicator		Stage	Weight	Total Weight
15	Years in cooperation receiving public funding	Minimum qualification requirements	12.50%	100%
16	Number of projects executed	Minimum qualification requirements	12.50%	
17	Volume of funds managed	Minimum qualification requirements	12.50%	
18	Audit of annual accounts	Minimum qualification requirements	12.50%	
19	Volume of profits or losses	Minimum qualification requirements	12.50%	
20	Evolution of shareholders' equity	Minimum qualification requirements	12.50%	
21	Number of staff hired	Minimum qualification requirements	12.50%	
22	Social base: Number of members and private donors.	Minimum qualification requirements	12.50%	

Source: Own elaboration.

For the purposes of this research, in order to weight these indicators, and given that the agency gives 100% weight to compliance with each of the associated requirements, it will be considered that together they weigh 100% for the agency, so that each of them will have a weight in this study of 1/8 of 100%, i.e., 12.5%.

Table 4. Indicators extracted from the quantitative and qualitative assessment of the AECID qualification.

Indicator		Stage	Weight (Real of the Agency)	Total Weight
23	Solvency Ratio	Scale of qualification	6.00%	49.35/100 The rest of the indicators up to 100 points cannot be assessed due to a lack of public data or because they are formal (for example, the number of women in managerial positions or presence in priority countries for Spanish cooperation).
24	Average Workforce	Scale of qualification	2.25%	
25	Temporariness Rate (average percentage of permanent contracts with respect to the average workforce).	Scale of qualification	2.00%	
26	Average ratio of cooperating partners per project (includes the AECID requirement: number of development workers in the field)	Scale of qualification	3.15%	
27	Number of partners	Scale of qualification	3.00%	

Table 4. Cont.

	Indicator	Stage	Weight (Real of the Agency)	Total Weight
28	Number of volunteers Includes the following AECID indicators: (1) Number of volunteers (with a weight of 3 points out of 100) and (2) The organisation has had volunteer personnel continuously for the last five years and has complied with Law 6/1996 and Law 23/1998 (BOE 1998) (which is valued at 1.08 points out of 100).	Scale of qualification	4.08%	49.35/100 The rest of the indicators up to 100 points cannot be assessed due to a lack of public data or because they are formal (for example, the number of women in managerial positions or presence in priority countries for Spanish cooperation).
29	Volume of funds earmarked for cooperation	Scale of qualification	2.00%	
30	Years of dedication of the NGDO to development cooperation	Scale of qualification	4.00%	
31	Number of countries in which it is present	Scale of qualification	2.00%	
32	Publicity on the website of Strategic Planning. It covers the following AECID requirements: (1) The existence of a document setting out the mission, vision, and values (scored 0.45 points out of 100); (2) The publicity of this document (scored 0.45 points out of 100); and (3) Staff, including volunteers, are aware of the mission, vision, and values (scored 0.45 points out of 100).	Scale of qualification	4.05%	
33	The names of the people who make up the board of trustees or the board of directors are published with a brief biographical sketch or curriculum vitae on the website. It meets AECID requirements: (1) The members of the governing bodies are publicised (scored 0.9 points out of 100) and (2) The members of the governing bodies are renewed (scored 0.45 points out of 100).	Scale of qualification	1.35%	
34	It has an updated website. It includes the AECID indicators: (1) The NGDO has an updated website (rated 1.5 points out of 100) and (2) There is a website that complies with the provisions of Law 19/2013 (rated 2.475 points out of 100).	Scale of qualification	3.98%	
35	Organisational chart and document that defines the functions of each department and job position.	Scale of qualification	0.50%	
36	Publication of the annual activity report on the website. This indicator includes the following AECID indicators: (1) The governing body annually monitors compliance with the objectives established in the strategic planning (rated 0.9 points out of 100); (2) The entity's activities are in line with the vision and strategic plan (rated 0.9 points out of 100); (3) There is a document that includes the annual planning of activities, approved by the governing body (rated 1.6 points out of 100); (4) The governing body is aware of the degree of compliance with the objectives of the annual planning (rated 1 point out of 100); (5) The governing body is aware of the degree of compliance with the objectives of the annual planning (rated 1 point out of 100). (rated 1.6 points out of 100); (6) The governing body is aware of the degree of compliance with the objectives of the annual planning (rated 1 point out of 100); and (7) The activities report is sent annually to members and collaborators (rated 0.9 points out of 100).	Scale of qualification	5.30%	

Table 4. Cont.

Indicator		Stage	Weight (Real of the Agency)	Total Weight
37	Membership in networks and platforms. This includes the following AECID indicators: (1) The NGDO belongs to a specific Spanish platform for Development Cooperation (rated 1.05 points out of 100) and (2) The NGDO belongs to an international platform and actively participates (rated 1.05 points out of 100).	Scale of qualification	2.10%	49.35/100 The rest of the indicators up to 100 points cannot be assessed due to a lack of public data or because they are formal (for example, the number of women in managerial positions or presence in priority countries for Spanish cooperation).
38	The website contains the investment policy.	Scale of qualification	0.90%	
39	Profit or loss.	Scale of qualification	0.315%	
40	The audit reports are published on the website. It includes the following AECID indicators: (1) The audit reports of the last three years do not reflect uncertainty regarding the continuity of the activity (valued with 0.525 points out of 100), and (2) In the last six years, the audit reports of the annual accounts do not present the same qualification on two or more occasions (valued with 0.315 points out of 100).	Scale of qualification	0.84%	
41	Evolution of equity	Scale of qualification	0.315%	
42	The annual accounts are published on the website. Includes the AECID indicator: access to accounting records.	Scale of qualification	0.315%	
43	Percentage of public funding This includes AECID's requirement that in the last five years, considered separately, more than 80% of income does not come from public subsidies.	Scale of qualification	0.90%	

Source: Own elaboration.

Once the first phase has been completed, the agency carries out a quantitative and qualitative assessment of the entity. The criteria and indicators extracted from the regulations that can be verified with the published data are shown in Table 4. It should be noted that those of the indicators established by the agency that are closely related to each other have been grouped together in the same indicator, and that in other cases the wording of the indicator has been adapted so that it can be verified with public data; these points are indicated in Table 4.

Analysing all the indicators derived from the regulations that set out the AECID's eligibility criteria, the following conclusions can be drawn: Firstly, that there are repeated indicators or indicators that value similar concepts, which have been grouped together in a single indicator, adding the weighting given to each of them. Thus, for example, Table 5 shows how the number of workers is valued in indicators 3, 13, 22, and 25, which can be grouped together in a single indicator, called average workforce, with a weighting that is the result of the sum of the weights of each indicator grouped together. The average workforce would be an indicator with a total weight of 30.11 percentage points out of 260.78 possible.

Secondly, it is concluded that the indicators can be summarised, by similarity, in four criteria: experience, transparency and internal functioning, human resources, and financial resources. Table 6 summarises the indicators extracted from the AECID's eligibility criteria, classified into criteria, and summarised under a single criterion for all indicators that are similar. For each indicator, those to which it would correspond for the agency are indicated.

Table 5. Grouping of indicators corresponding to the average workforce indicator.

Indicator		Stage	Weight
3	Number of workers employed.	AECID Registration (2)	12.50%
12	Number of workers according to their funds	Scale grant (4)	2.86%
21	Number of staff employed	Minimum qualification requirements (3.2.1)	12.50%
24	Average number of staff	Qualification scale (3.2.2)	2.25%
Average Workforce			30.11%

Source: Own elaboration.

Table 6. Aggregate indicators and standardised weights used by AECID.

Summary Indicator		Indicators	Total Weight	Standardised Weight
Experience	Volume of funds managed	17 and 29	14.50%	5.50%
	Number of projects implemented	16	12.50%	5%
	Number of countries in which it is present	31	2%	0.70%
	Age of the NGDO	15 and 30	16.50%	6%
Transparency and performance	Publication of the organisation chart on the website	1 and 34	13.16%	5%
	Organisational chart and a document that defines the functions of each department and job position.	2	0.50%	0.20%
	Publication of the strategic planning on the website.	32 and 34	4.71%	2%
	Publication of the annual activity report on the website.	8, 9, 34 and 36	19.41%	7.50%
	Have the annual accounts and audit reports published on the website.	10, 18, 34, 40 and 42	15.27%	6%
	Having the names of the people who make up the board of trustees or the board of directors with a brief biographical sketch or curriculum vitae published on the website.	11, 33 and 34	2.96%	1%
	The website contains the investment policy.	34 and 38	1.56%	0.60%
	Membership in networks and platforms.	14 and 37	4.96%	2%
	Average workforce	3, 12, 21 and 24	30.11%	11.50%
Human Resources	Temporary staffing rate	25	2%	0.70%
	Number of volunteers	4 and 28	16.58%	6.50%
	Number of development workers (average ratio of development workers per project).	5 and 26	15.65%	6%
	Social base: number of partners and private donors.	2, 22 and 27	28.00%	11%
Financial Resources	Volume of private funds received	6 and 13	13.93%	5%
	Volume of public donations received	7 and 13	13.93%	5%
	Percentage of public funding	43	0.90%	0.30%
	Solvency ratio (current assets/current liabilities)	23	6.00%	2.50%
	Evolution of shareholders' equity	20 and 41	12.82%	5%
	Profit or loss	19 and 39	12.81%	5%
Total Weight			260.78%	100

Source: Own elaboration.

3.2. Analysis of the European Union's Eligibility Criteria (Europe-Aid)

The European Union has been engaged in development cooperation since its creation as the European Economic Community in 1957. EU development cooperation policy is managed by the Commission's Directorate-General for International Cooperation and Development (DG DEVCO), which since 2015 has brought together the former Europe-Aid

Cooperation Office (AIDCO, created in 2011) and the Directorate-General for Development and Relations with ACP (African, Caribbean, and Pacific) States.¹

The EU accompanies each call for proposals with a guide for applicants, which describes, among other things, the purpose of the call for proposals, the eligibility requirements for applicants, and the evaluation criteria for the selection of competing NGOs and the award of the grant. In general, in order to be eligible to tender for grants, the EU requires NGOs to meet certain requirements related to: (i) the nationality of the entities (contracts are only concluded and grants awarded to entities from certain countries) or (ii) to the origin of the goods they intend to purchase with EU funds. In addition, it asks entities to provide a declaration of honour certifying that they have not committed any irregularities in their previous operation. These irregularities have to do with bankruptcy proceedings, fraud, corruption, or terrorist offences in previous contracts with the EU. All these requirements cannot be considered indicators of good performance. In general, in order to check the financial and operational capacity of NGOs, the EU requires applicants to be registered in the Potential Applicant Data On-Line Registration (PADOR). Organisations wishing to obtain EU funds must provide PADOR with data on the organisation's profile, values, target population, sectorial and geographical experience, as well as financial data, human resources, and the composition of the governing body. However, the regulations do not make explicit to what extent and how these data are used (Europe-Aid Online Registration Service 2008).

Once registered in PADOR, an evaluation committee carries out a first analysis of the technical formulation of the project submitted by the NGO to obtain the grant. In this first analysis (called the analysis of the synthesis formulation), only aspects of the project are rated according to criteria that are individually assessed on a scale of 1 to 5 (1 would be very poor compliance and 5 would be very satisfactory compliance). Only the summary documents that have obtained a certain minimum total score will be taken into account for pre-selection purposes. These criteria, as they refer to specific aspects of the project submitted by the NGO for the grant (relevance and design), do not provide information on the general functioning of the organisation but only on how it will implement this particular project. Therefore, they cannot be considered as indicators of the organisation's performance.

After the pre-selection, in the next step, the EU reassesses, through the data provided in a 'complete' application form, aspects related to the specific action to be carried out. In this step the goal is reassessing its relevance and also assessing its effectiveness, viability, and sustainability, devoting a single criterion, to which it awards a weighting of 20 points out of 100, to the assessment of the operational and financial capacity of the applicants.

The indicators established by the EU linked to this criterion and their weighting are listed in Table 7. The scoring scale for each criterion ranges from 1 to 5, where 1 = very poor; 2 = poor; 3 = acceptable; 4 = satisfactory; 5 = very satisfactory. The NGO must score a minimum of 12 points on this criterion to pass the selection. The description of the indicators is ambiguous as terms such as "sufficient experience", "sufficient capacity", or "sufficient sources of funding" are used, so the final score given to each indicator will depend on the experts.

After the evaluation of the financial and operational capacity, the NGOs will be analysed by the evaluation committee by comparing the data declared to PADOR with the information contained in the statutes of the entity. These statutes include the financial reports, the audit report of the last year available, a legal entity form (showing the legal form of the entity and its personal data), a financial entity form (showing the bank in which the grant will be deposited), and a declaration of the entity stating that it has the means to carry out the action and commits itself to do so.

In no case does the EU provide information on the indicators and the requirements associated with them that serve as a basis for the assessment of NGOs. It is not possible to know on what basis the EU finally decides which organisations to subsidise out of all those that have passed the previous stages analysed.²

Table 7. Indicators for the financial and operational capacity criterion of the EU full application form.

Section	Maximun Score
1. Financial and operational capacity	20
1.1 Do the applicants and, if applicable, their affiliated entities have sufficient project management experience?	5
1.2 Do the applicants and, if applicable, their affiliated entities have sufficient technical expertise? (in particular, knowledge of the issues to be addressed)?	5
1.3 Do the applicants and, if applicable, their affiliated entities have sufficient management capacity (in particular staff, equipment, and capacity to manage the budget of the action)?	5
1.4 Does the lead applicant have stable and sufficient sources of funding?	5

Source: Assessment table of the completed application form. (Comisión Europea 2016).

The analysis of the data requested by the EU has made it possible to identify possible indicators that have been classified into four criteria (see Table 8). As the financial and operational capacity criterion has a weighting of 20% for the EU in the subsidy scale, it has been divided among all the indicators identified in this study, giving them all the same weighting.

Table 8. Indicators derived from the EU’s financial and operational capacity criterion.

Criterion	Indicator	Weight	Standardised Weight
Project management experience	Number of projects carried out	1.67%	8.5%
	Number of countries in which it operates	1.67%	8.5%
	Volume of funds implemented	1.67%	8.5%
Technical expertise	Number of sectors in which the NGDO operates	2.50%	12.5%
	Years of experience	2.50%	12.5%
Management capacity	Number of staff at headquarters/average number of staff	1.25%	6%
	Number of development workers/ratio of development workers per project.	1.25%	6%
	Average number of volunteers	1.25%	6%
	Volume of non-current assets	1.25%	6%
Stable and sufficient sources of funding	Number of private donors	1.67%	8.5%
	Number of public donors	1.67%	8.5%
	Percentage of public funding	1.67%	8.5%
Total Weight		20	100

Source: Own elaboration.

3.3. Analysis of USAID Eligibility Criteria

The United States regulates the awarding of development cooperation grants in the Foreign Assistance Act of 1961 and in Regulation 2 CFR 200: Uniform Administrative Requirements, Cost Principles, and Audit Requirements for Federal Awards (U.S. Government Publishing Office [GPO], n.d.). USAID, as the agency in charge of US foreign policy in the area of development cooperation, works under a series of rules, directives, encompassed in a system called ADS (Automated Directives System).

When USAID issues a Notices of Funding Opportunity (NOFO), any NGDO that wants to apply for a USAID grant must have a DUNS (Data Universal Numbering System) number, which identifies the organisation. It also must have a NATO Commercial and Government Entity Code (N Cage) and be registered in the United States Government Grants Management System (SAM) (US Federal Register of Contractors, USFCR 2020).

Obtaining the DUNS and (N) Cage does not require any additional requirements other than providing the organisation’s identification data.

Registration with SAM also involves providing information on the organisation’s age, fiscal year-end date, total revenues for the last three years, number of employees for the last year, compliance with environmental laws, tax obligations, and federal antitrust and corruption laws (affecting the organisation’s officers and directors). These include that the organisation’s representatives and officers must not be debarred from working with the federal government or have been convicted of antitrust violations, compliance with previous contracts with the US government, declaration of non-procurement of goods and services from countries that do not guarantee labour rights or have restricted foreign policies³, and compliance with US laws on non-discrimination on the basis of race, sex, religion, colour, sexual orientation, or national origin. Upon registration with SAM, organisations seeking to receive a USAID grant must sign a declaration of honour⁴ certifying compliance with certain legal and ethical obligations.

Once these pre-requisites for registration have been met, USAID subjects the eligibility of organisations to a four-stage screening process.

The first phase consists of passing general criteria, usually related to the NGDO’s experience, resources, and compliance with US legislation. The second phase involves the fulfilment of the criteria set out in the specific call for proposals for a grant. These usually refer to the applicant organisation’s previous experience both sectorially and geographically, taking as indicators the number of successful actions previously carried out in the same sector and the number of years working in the country. Once the two previous phases have been completed, USAID submits the NGDO to an initial review. The criteria for passing this initial evaluation are established in each call for proposals, although, in general, they are in line with those listed in Table 9.

Table 9. USAID Initial Review Assessment Criteria.

Technical Aspects (65 points)
Creative solutions for the priority intervention areas (30%)
Coherent implementation plan and likelihood of implementation within the proposed timeframe (13%)
Describe any potential risks of the proposed activity to (i) the health and safety of participants and beneficiaries, or to (ii) the environment and local ecosystems; and detail risk mitigation plans (2%)
Proposed monitoring and evaluation factors as indicators to measure programme impacts (10%)
Strategy for capturing knowledge and sharing lessons learnt (10%)
Management and institutional capacity (35 points)
Human resources (staff, partners, and/or consultants) to implement all project components, including technical, administrative, financial, and monitoring and evaluation. (25%)
Previous performance experience of the applicant in similar project activities (10%)

Source: USAID (2018).

USAID does not establish the specific indicators that will be studied to verify compliance with each of the criteria, nor the weight to be given to each indicator. Thus, for example, one of the criteria that appears is previous experience in activities similar to those of the project with a weight of 10%. The fact is that it does not say whether experience will be measured by the number of projects previously carried out in the same sector, by the number of projects carried out in the same country, or whether other indicators will be taken into account. Nor does it indicate what the requirements are for each indicator to be considered positive or what weight each indicator has when assessing the criterion in question if there is more than one indicator to check the same criterion. Furthermore, USAID gives the scale with the weights associated with each criterion but does not establish

the minimum score to be achieved in order to pass this preliminary phase. The only thing it adds is that the applicant will be informed in writing after the initial review.

In relation to the criteria set out in Table 9, those referring to the technical approach of the project (technical aspects) will not be taken into account in this study. The reason is that they are criteria related to a specific project in time and therefore represent a declaration of intent, but do not demonstrate the real functioning of the entity.

The criteria set out in the management and organisational analysis (management and institutional capacity) are considered, but their indicators are not provided by the agency. It is known that the data on which the decision is based are contained in SAM and in the documents requested from the entity. In consequence, the analysis of previous experience will be based on the data on the seniority of the organisation contained in SAM and on the applicant’s performance history. It is not possible to know from the data in the performance history, which specific indicators are used to verify this experience, which has a weighting of 10% in the grant analysed. In this work, two indicators are considered: the length of service of the organisation and the number of projects previously carried out, both with a weight of 5% each (see Table 10).

Table 10. Indicators derived from the USAID targeting phases.

Phase	Criterion	Indicators	Weight	Total Weight
1. To work with USAID	Experience	Age of the NGDO	33.33%	100%
	Human resources	Number of employees	33.33%	
	Financial resources	Income in the last three years	33.33%	
2. As required by the call for proposals	Experience	Number of countries in which it operates	50.00%	100%
		Number of projects carried out	50.00%	
3. Initial review	Experience	Age of the NGDO	5.00%	35%
		Number of projects carried out	5.00%	
	Human resources	Number of employees in the last year	25.00%	

Source: Own elaboration.

To check human resources, USAID uses the indicator number of employees in the last year. Although, as the criterion is worded, human resources—staff, partners, and/or consultants—to implement all project components, including technical, administrative, financial, and monitoring and evaluation, seems to only analyse human resources. Considering that total income for the last three years is also requested (in SAM) and that audit reports are required (in the grant application), it is concluded that the analysis also covers financial resources. However, it is not known what data is analysed from these documents and therefore which indicators are associated with this requirement, so we will only consider the USAID explicit indicator of the number of people employed in the last year, with an associated weight of 25% (see Table 10).

If the NGDO passes the initial assessment and the three previous phases, it will go on to a new, more detailed analysis of its capacity, generally by means of a “pre-assessment survey” (NUPAS—Non-U.S. Organisation Pre-Award Survey). The NUPAS involves the analysis and evaluation, by a group of experts appointed by the agency, of 29 criteria, classified into six categories or blocks of equal weight, i.e., each of the blocks has a weight of 1/6 of the total score.

To verify compliance, each criterion may have one or more indicators, which will be evaluated according to a scale ranging from a score of 1 (inadequate compliance) to 4 (strong compliance). The score of the corresponding criterion is calculated as the average of the score obtained for each indicator. The block score is calculated as the average of the scores for each criterion and the overall score, the score obtained by the organisation, as the average of the scores in each of the 6 blocks. Depending on the score obtained, the experts conducting the evaluation will issue an audit report with three possible results: (1) totally

positive, the organisation is eligible to receive a grant; (2) positive, but with caveats, the organisation has operational deficiencies in certain aspects that USAID considers key and must be corrected to be eligible; and (3) negative, the organisation does not meet the minimum operational requirements to ensure proper management of USAID funds and the implementation of the activities to be funded. It is not known what average score the organisation must obtain for the report to be of one type or the other.

It should be noted that the NUPAS in some cases specifies the indicators of the criterion (for example, the solvency ratio as an indicator of the absorption capacity criterion), but in several cases, compliance with the criterion or the associated indicator is defined generically, using as a source of verification the observation by a group of experts of the organisation's internal procedures. This is completed without specifying which variables are used by the group of experts or what must be included in these procedures for the indicator to be considered positive. Table 11 shows those indicators specified by the NUPAS, taking into account that those that the experts verify by means of data collected through observation have not been considered.

Table 11. USAID eligibility criteria, indicators, and weights.

Phase	Criterion	Indicators	Weight	Total Weight
1. To work with USAID	Experience	Age of the ONGD	33.33%	100%
	Human resources	Number of employees	33.33%	
	Financial resources	Revenues in the last three years	33.33%	
2. As required by the call for proposals	Experience	Number of countries in which it operates	50.00%	100%
		Number of projects carried out	50.00%	
3. Initial review	Experience	Age of the ONGD	5.00%	35%
		Number of projects carried out	5.00%	
	Human resources	Number of employees in the last year	25.00%	
4. NUPAS	1.3 Organisational structure	The organisation chart of the organisation can be found on its website	1.67%	26.42%
		The organisation chart shows the lines of authority, responsibility, and communication.	1.67%	
	1.4 Governance	Disclosure of data and transparency	3.33%	
	1.5 Internal Control	The statutes are available on the organisation’s website	1.67%	
		The statutes contain the functions of the board, a term limit for board members, and a system for the renewal of board members.	1.67%	
	2.9 Financial resource management 2.12 Audit and review of financial statements	Annual accounts and audit reports are published on the website.	2.60%	
	2.10 Sources of funding	Number of public donors	0.64%	
		Number of private donors	0.64%	
	4.1 Human resources policy	The organisation chart of the organisation is published on the website.	2.10%	
		The organisation chart shows the functions of each department and employee.	2.10%	
	6.1 Treasury management	Solvency ratio	8.33%	

Source: Own elaboration.

Having analysed the criteria, indicators, requirements, weightings, and sources of verification used by the Spanish, European, and American agencies, in the following section we will deduce, by comparing them, which criteria and indicators can be used to verify the internal performance of NGDOs.

3.4. Choice of Criteria and Indicators to Study the Good Performance of NGDOs

Having analysed the eligibility criteria of the Spanish, European, and American agencies, it only remains to deduce, by comparison, which criteria are used to verify the good internal performance of the entities, the indicators and requirements associated with them, and the weights of these indicators.

The indicators obtained from the three donors analysed and the standardised weights calculated for each indicator (Tables 6, 8 and 11) have been summarised in a single Table 12. The weights of each agency for each indicator have been rounded, and the indicators not covered by the different agencies are listed as not applicable (N.A.).

From the study carried out in the previous section, we conclude that the criteria most frequently used by donors are four: (1) experience; (2) transparency and internal organisation or functioning; (3) human resources; and (4) resources and financial management. We now classify these indicators into the criteria listed above, establishing for each of them its realisation value or ideal value from the donor's point of view, in accordance with the regulations analysed in the previous section.

The first criterion is that of experience, collected by the agencies by means of the following indicators: 1. number of projects carried out previously; 2. length of service of the NGDO, which includes indicators related to the number of years the organisation has been involved in cooperation; 3. volume of funds managed; and 4. number of countries in which it operates.

For donors, experience is a criterion to be maximised, awarding more points in the grant access scales to those organisations that offer higher numbers in different criteria. These are the number of projects executed, volume of funds managed, countries in which they operate, or the organisation's seniority, which is logical, as the greater the experience, the better the internal functioning of the organisation due to the lessons learnt.

The second criterion is transparency and internal organisation. The assessment of the internal functioning of the organisation is usually exhaustive by the three donors analysed. From the point of view of transparency, the Spanish and American agencies assess whether the entity has a website or electronic headquarters. In the Spanish case, it is also made explicit that in order for the entity to be considered transparent, the webpage must include "activity reports, audited accounts, the composition of its governing bodies and technical teams" (AECID 2019, article 11.2). In the case of USAID, it only refers to transparency within the criterion "governance and practice of good corporate governance", assessing "disclosure of data and transparency" but without defining what this consists of (NUPAS n.d.).

On the other hand, if only the documents used by the three donors are taken into account to analyse the internal functioning of the organisation, without considering those that are used by a specific donor (such as the investment policy requested by AECID-AECID 2017-), it can be deduced that the internal functioning is based on the analysis of different documents. These are the statutes (where it is also possible to assess whether the organisation has the functions of the board members, a procedure for the renewal of these members, and a term limit), the strategic planning (which also includes the organisation's mission, vision, and values), the activities report (which also allows us to know whether the actions carried out and the annual planning are monitored), the organisation chart (which, if complete, will also allow us to know whether the organisation's lines of responsibility, authority, and communication are delimited), the annual accounts, and the audit reports.

For donors, transparency and internal functioning is a maximisation criterion, with higher scores being awarded to those NGDOs that achieve higher scores on these indicators.

The third criterion covers human resources. Human resources are considered to be those people who work for the organisation, with or without an employment contract.

Table 12. Comparison of AECID, Europe-Aid, and USAID Eligibility Indicators.

Indicator	Aecid Weight	Europe-Aid Weight	Usaid Weight
Volume of funds managed	5.5%	8.5%	N.A.
No. of projects implemented	5%	8.5%	21%
Number of countries in which it is present	0.7%	8.5%	19%
Number of sectors in which the NGDO is active	N.A.	12.5%	N.A.
Age of the NGDO	6%	12.5%	14.5%
Publication of the organisation chart on the website.	5%	N.A.	2%
The organisation chart delimits the functions of each department and job position.	0.2%	N.A.	2%
Publication of the annual report of activities on the website.	7.5%	N.A.	
The annual accounts and audit reports are published on the website.	6%	N.A.	1%
The names of the people who make up the board of trustees or the board of directors with a brief biographical sketch or curriculum vitae are published on the website.	1%	N.A.	N.A.
The statutes are published on the organisation's website.	N.A.	N.A.	1%
The statutes contain the functions of the board, a term limit for board members, and a procedure for their appointment and dismissal.	N.A.	N.A.	1%
Membership in networks and platforms.			
Membership in networks and platforms.	2%	N.A.	N.A.
Number of employees/average workforce	11.5%	6%	22%
Temporariness rate	0.7%	N.A.	N.A.
Number of volunteers	6.5%	6%	N.A.
Number of development workers/average ratio of development workers per project	6%	6%	N.A.
Publicity on the website of strategic planning.	2%	N.A.	N.A.
The website includes the investment policy.	0.6%	N.A.	N.A.
Social base: number of partners and private donors.	11%	8.5%	0.25%
Number of public donors		8.5%	0.25%
Volume of private funds received	5%	N.A.	N.A.
Volume of public donations received	5%	N.A.	N.A.
Percentage of public funding	0.3%	8.5%	N.A.
Solvency ratio (current assets/current liabilities)	2.5%	N.A.	3%
Development of shareholders' equity	5%	N.A.	N.A.
Volume of non-current assets	N.A.	6%	N.A.
Income in the last three years	N.A.	N.A.	13%
Profit or loss.	5%	N.A.	N.A.
	100%	100%	100%

Source: Own elaboration.

Another indicator that is considered interesting and which is only assessed by the AECID is the temporality rate, calculated, as defined by the AECID (2017), as a percentage of permanent contracts over the total staff. However, this definition does not correspond to that usually used in academic literature or by official⁵ bodies, which define the temporality rate as the ratio of temporary contracts over the total. This rate makes it possible to evaluate whether the organisation has a stable workforce retaining their learning and experience and

also assesses whether the staff turnover is low, which can denote good staff management (good training, good remuneration, transparency in decisions, etc.).

The last criterion is financial resources. As with the criterion of transparency and internal organisation, the organisation’s financial resources are assessed by donors using indicators and requirements that are not clearly publicised. For this, there is a group of people who are experts for the donor who analyse the organisation’s policies, internal functioning, and financial documents by observation.

However, from the comparison of the three agencies analysed, it can be deduced that the most important thing for the donor is to verify the existence of sufficient resources in the organisation to carry out the subsidised project and the stability of the organisation’s financial resources.

Table 13 lists the indicators that have been deduced to analyse the performance of the institutions. We indicate whether for this group it is an indicator to be maximised or minimised, the indicators that have been discarded, the weights attributed to each indicator by each agency, and the total weight that this indicator has for the donors. Indicators not considered by the different agencies are shown as not applicable (NA).

The indicators discarded are due to the fact that only one donor has taken them into account. The reason is that they are not considered to be an effective indicator for analysing the performance of the entities, such as the publication of the investment policy on its website, the number of sectors in which the NGDO operates, or the number of networks to which it belongs.

As far as profits are concerned, in a non-profit sector, not making profits is not considered a factor to be taken into account when analysing efficiency. The distribution of profits is limited by law (Article 3 of Law 49/2002 of 23 December 2002 on the tax regime for non-profit organisations and tax incentives for patronage), so that making a profit cannot be the reason for collaborating with the organisation or the objective expected of it.

Table 13. Criteria, indicators, and eligibility weights for public donors.

Summary Indicator		Max/Min	AECID Weight	Europe-Aid Weight	Usaid Weight	Total Weight	Normalised Weight
Experience	No. of projects implemented	Max.	5.0%	8.5%	21.0%	35%	13.0%
	Age of the NGDO		6.0%	12.5%	14.5%	33%	12.0%
	Volume of funds managed		5.5%	8.0%	N.A.	14%	5.0%
	Number of countries in which it is present		0.7%	8.5%	19.0%	28%	10.0%
Human Resources	Number of employees/average staff	Max.	11.5%	6.0%	22.0%	40%	14.5%
	Number of volunteers		6.5%	6.0%	N.A.	13%	5.0%
	Number of development workers/average ratio of development workers per project		6.0%	6.0%	N.A.	12%	4.0%
	Temporary staff ratio		0.7%	N.A.	N.A.	1%	0.3%
Financial Resources	Social base: number of partners and private donors.	Max.	11.0%	8.5%	0.25%	20%	7.0%
	Number of public donors	Max.	N.A.	8.5%	0.25%	9%	3.0%
	Volume of private funds received	Max.	5.0%	N.A.	13.0%	11.5%	4.0%
	Volume of public donations received	Max.	5.0%	N.A.		11.5%	4.0%
	Percentage of public funding	Min.	0.30%	8.5%		9%	3.0%
	Solvency ratio (current assets / current liabilities)	Max.	2.5%	N.A.	3.0%	6%	2.0%
	Development of own funds	Max.	5.0%	N.A.	N.A.	5%	2.0%

Source: Own elaboration.

Table 13. Cont.

Summary Indicator			Max/Min	AECID Weight	Europe-Aid Weight	Usaid Weight	Total Weight	Normalised Weight
Transparency and Internal Functioning	Details of members of governing bodies	To have on the website the names of the people who make up the Board of Trustees or the Board of Directors with a brief biographical sketch or curriculum vitae.	Max.	1.0%	N.A.	N.A.	1%	0.4%
	Statutes	The statutes are available on the website of the organisation.		N.A.	N.A.	1.0%	1%	0.4%
		The statutes include the functions of the board, a term limit for its members, and a procedure for their appointment and dismissal.		N.A.	N.A.	1.0%	1%	0.4%
	Organisation chart	The organisation chart is available on the website.		5.0%	N.A.	2.0%	7%	2.5%
		The organisation chart delimits the functions of each department and job.		0.20%	N.A.	2.0%	2%	1.0%
	Strategic planning	Strategic planning is published on the website		2.0%	N.A.	N.A.	2%	1.0%
Annual report of activities	Publication of the annual activity report on the website.	7.5%	N.A.	N.A.	8%	3.0%		
Annual accounts and audit reports	The annual accounts and audit reports are published on the website.	6.0%	N.A.	1.0%	7%	2.5%		
Total Weights Assumed			92.4%	81.0%	100%	273.4%	100%	
Indicators not considered:			N.A.	N.A.	N.A.			
Membership in networks and platforms.			2.0%	N.A.	N.A.			
Number of sectors in which the NGDO operates.			N.A.	12.5%	N.A.			
The website contains the investment policy.			0.60%	N.A.	N.A.			
Volume of non-current assets.			N.A.	6.0%	N.A.			
Profit or loss.			5.0%	N.A.	N.A.			
Weights Indicators Discarded			7.60%	18.50%	0.00%			
Total			100%	100%	100%			

Source: Own elaboration.

4. Empirical Results: Promethee Approach

The efficiency of the 74 NGDOs that form part of the State NGDO Coordinating Committee is analysed using the indicators obtained from the donor eligibility criteria listed in previous tables. The data were obtained from those published by the Coordinating Committee in its recent sector report (Coordinadora de ONG para el Desarrollo-España 2017), from the reports that NGDOs send to the Coordinating Committee and that the latter publishes on its website (Coordinadora de ONGD-España), and from the data on the website of each of the NGDOs in the sample. For the application of the method, the free software Visual Promethee 1.4 Academic Edition (available at <http://www.prometheegaia.net/software.html>; accessed on 2 June 2019) will be used. The steps are as follows: (i) Configuration of the multi-criteria decision matrix; (ii) The different alternatives (the 74 NGDOs) and the indicators (listed in previous tables) that will be used to determine the efficiency of each NGDO, their values and weights, are loaded into the programme. Each

indicator is assigned its unit of measurement, and it is indicated whether it is an indicator to be maximised (the higher the indicator, the higher the value) or minimised (the lower the value, the higher the value); (iii) Two groups are created, the outputs (represented in the study by a green circle) and the inputs (represented in the study by a blue square), classifying each indicator in the corresponding group. Table 14 shows the NGDOs ranked according to their efficiency and the grants (in euros per project -€/pyto-) they have received from Spanish public donors during the analysed period, obtained from the BDNS. It also shows the order of the NGDOs from the highest to the lowest volume of grants received (in €/per project).

Table 14. Net flows of inputs and outputs.

NGDO	Input Virtual (X)	Output Virtual (Y)	Effectives	Efficients	Frugals	Inefficients
			Inputs + Outputs +	Inputs – Outputs +	Inputs – Outputs –	Inputs + Outputs –
AV	0.4095	0.0382	YES			
ANC	0.3442	0.1354	YES			
CERAI	0.0086	0.0044	YES			
CIDEAL	0.0544	0.0651	YES			
JCONGD	0.2182	0.1554	YES			
PROYDE	0.1066	0.3573	YES			
SED	0.022	0.2273	YES			
ACF-E	–0.7159	0.2991		YES		
ALBOAN	–0.3279	0.3587		YES		
ACPP	–0.2406	0.3075		YES		
EP/HE	–0.143	0.0712		YES		
AeA	–0.5735	0.4165		YES		
Caritas Española	–0.355	0.6684		YES		
CESAL	–0.2384	0.3481		YES		
CODESPA	–0.1606	0.2432		YES		
C.R.E.	–0.7549	0.625		YES		
Educo	–0.415	0.3104		YES		
FM	–0.3557	0.1264		YES		
ANESVAD	–0.1943	0.4966		YES		
EC	–0.4578	0.3609		YES		
InteRed	–0.2624	0.0754		YES		
JYD	–0.0233	0.2878		YES		
Manos Unidas	–0.6079	0.6531		YES		
MdM	–0.5899	0.2372		YES		
FAMME	–0.5416	0.5424		YES		
MPDL	–0.3521	0.3004		YES		
Mundubat	–0.1355	0.0701		YES		
Oxfam Intermón	–0.882	0.6327		YES		
SC	–0.6777	0.3603		YES		

Table 14. Cont.

NGDO	Input Virtual (X)	Output Virtual (Y)	Effectives	Efficients	Frugals	Inefficients
			Inputs + Outputs +	Inputs – Outputs +	Inputs – Outputs –	Inputs + Outputs –
ADRA	−0.1991	−0.1411			YES	
APS	−0.4853	−0.0641			YES	
AdTE	−0.0545	−0.119			YES	
ASF	−0.1696	−0.2238			YES	
Tdh E	−0.0393	−0.2148			YES	
H+D	−0.0917	−0.0191			YES	
MZC	−0.1097	−0.1116			YES	
ONGAWA	−0.2461	−0.071			YES	
PcD	−0.0174	−0.1095			YES	
Plan Internacional	−0.5075	−0.0716			YES	
SETEM	−0.3336	−0.1425			YES	
VSF	−0.1131	−0.0391			YES	
AIDA	0.2014	−0.0432				YES
AIETI	0.2233	−0.0669				YES
AMREF	0.3131	−0.1551				YES
FONTILLES	0.0419	−0.0229				YES
CGJP	0.6827	−0.1343				YES
COOPERACCIO	0.2602	−0.1363				YES
CI ONG	0.0182	−0.1793				YES
Esf	0.0654	−0.4028				YES
EDIFICANDO CN	0.6391	−0.1452				YES
FAD	0.311	−0.2141				YES
F.S.F.E.	0.3886	−0.0383				YES
FERE-CECA	0.3224	−0.1347				YES
FISC	0.1483	−0.4785				YES
Fundación Adsis	0.2722	−0.149				YES
FEA	0.4434	−0.4297				YES
FRS	0.3022	−0.3876				YES
FDV	0.3992	−0.1798				
FIE CIPIE	0.4898	−0.1718				YES
Madreselva	0.3852	−0.256				YES
Mainel	0.3572	−0.2376				YES
FUDEN	0.2768	−0.1634				YES
F1M	0.358	−0.5621				YES
PROCLADE	0.1331	−0.0536				YES
Promoción Social	0.2568	−0.0399				YES
IA	0.3279	−0.5557				YES

Table 14. Cont.

NGDO	Input Virtual (X)	Output Virtual (Y)	Effectives	Efficients	Frugals	Inefficients
			Inputs + Outputs +	Inputs – Outputs +	Inputs – Outputs –	Inputs + Outputs –
ISCOD	0.2179	–0.2432				YES
FMDLP	0.2931	–0.1467				YES
O.C.A.S.H.A.	0.4616	–0.0606				YES
PROSALUS	0.164	–0.1421				YES
PS	0.2603	–0.3319				YES
PPHH	0.6182	–0.4967				YES
Rescate	0.0663	–0.1403				YES
SOTERMUN	0.5095	–0.5484				YES
Total			7	22	12	33

Source: Own Elaboration Form Visual PROMETHEE 1.4 Academic Edition.

Taking the data to a spreadsheet and drawing a Cartesian plane where the virtual input is represented on the abscissa axis and the virtual output on the ordinate axis, Figure A1 (Appendix A) is obtained.

The efficient frontier has also been represented by joining the NGDOs that occupy the best position (Oxfam Intermón, Manos Unidas, and Cáritas Española), those with the highest net flow of output, given their inputs, or the lowest net flow of inputs given their outputs. It is important to take into account that we are working under the assumption that any positive convex linear combination of efficient NGDOs also belongs to the frontier, so the efficient frontier will be the one represented in Figure A1 by the blue line.

The virtue of the graphical representation is that it allows, in an intuitive way, to know what position each entity occupies, which entities are closer to it, and what is the distance that separates it from the frontier. Moreover, it also shows in which direction (if the frontier is closer to it by increasing the outputs, as would be the case of FAMME, or by decreasing the inputs, as would be the case of ACF-E).

By means of the graphical representation, NGDO managers can have a clear vision of which variables (inputs or outputs) they must act on to improve their position, and, by analysing the functioning of the NGDOs located closer to the frontier, they can approach the decisions to be taken to improve the results obtained with their management (Figure A1 in the Appendix A).

The results obtained show that 44.6% of the NGDOs (33 out of the 74 studied) operate inefficiently, compared to 29.7%, which are efficient. Moreover, 9.5% of the entities operate effectively, achieving their objectives, but at the cost of wasting resources, and 16.2% manage their resources well but do not manage to satisfy the interests of all their participants, taken as the objectives of the NGDOs in this work.

Considering only the achievement of objectives (efficient and effective entities), 39.2% of NGDOs are efficient on the output side. On the input side, 45.9% of the NGDOs have better management of their resources (efficient and frugal).

5. Conclusions

This paper aims to study the efficiency in the development cooperation sector, define the objectives and resources of the NGDOs, and find general indicators that allow verifying to what extent the entities are achieving their goals and using their resources to be able to apply them to an efficiency study.

The efficiency indicators are inferred from those used by public donors when they select the NGDOs to receive funding, the so-called eligibility criteria published in different norms and laws. Thus, efficiency is measured in this paper using the indicators collected

in the eligibility criteria of three of the main cooperation agencies: AECID, EuropeAid, and USAID.

Given this opacity and in order to know if the indicators published by donors were sufficient to determine the award of a public grant to a NGDO, a simulation of granting public funding was applied. For this, the 74 NGDOs that in 2016 (up to 2018–2023) belonged to the Spanish Coordinator were taken. With the donor variables, a study of the efficiency of the coordinator's NGDOs is carried out, using a technique that is an adaptation of the Promethee multi-criteria method known as PROMETHEE Productivity Analysis (PPA).

Our efficiency study reveals that 44.6% of the NGDOs (33 out of the 74 studied) operate inefficiently, compared to 29.7%, which are efficient. Moreover, 9.5% of the entities operate effectively, achieving their objectives, but at the cost of wasting resources, and 16.2% manage their resources well but do not manage to satisfy the interests of all their participants, taken as the objectives of the NGDOs in this work. Considering only the achievement of objectives (efficient and effective entities), 39.2% of NGDOs are efficient on the output side. On the input side, 45.9% of the NGDOs have better management of their resources (efficient and frugal).

Our results are important for academics, donors, and managers of these types of organisations because they must be efficient in the use of the money they receive. Our research has theoretical and practical implications. The different criteria, the indicators used to corroborate their fulfilment, and the sources from which donors draw the data to quantify the indicator are scattered across different standards without establishing a clear relationship between them. The weightings are provided in a general way for the criteria without disaggregating the weight of each indicator. The measure of compliance with the indicators is established in an imprecise manner with terms such as 'sufficient' or 'adequate', with a group of experts, appointed by the donor itself, subjectively assessing it by analysing the data provided by the entities or directly observing their functioning. Observation makes up for the absence of concrete, measurable, and measurable objective criteria.

Therefore, in this research and with the aim of finding out whether the indicators published by donors, were sufficient to determine the awarding of a public subsidy to an NGDO, a simulation of public funding was applied. Despite the opacity, the indicators inferred from donors are considered valid and sufficient, which is the main implication of our research. Valid in that they are objective and come from an important stakeholder group for NGDOs; sufficient because they allow us to assess all outputs, all objectives, and all inputs, all resources, of the organisations. Comparing the NGDOs that are efficient with those that receive the most funding from Spanish public donors shows that the most subsidised organisations are not precisely the most efficient, nor even the most effective.

As for the limitations of this research, one of them is the time period, due to the large amount of work involved in data collection, which is not systematised in any specific database. In relation to future research, our aim would be to use and compare the efficiency results with the Promethee technique and with TOPSIS in order to evaluate the possible results and make efficiency comparisons with both methods. Regulatory opacity hides other indicators and other motivations for granting public funding, which go beyond what is explicitly stated by donors. Moreover, the search for such indicators and motivations could be the subject of further studies.

Author Contributions: All authors contributed equally to this manuscript. All authors have read and agreed to the published version of this manuscript.

Funding: This research received no external funding.

Data Availability Statement: This research is supported by a database specifically constructed from data published by NGDOs. It is available upon request from potential readers.

Acknowledgments: We acknowledge the comments of the participants to the ACEDE Conference 2021.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

- ¹ Despite the creation of the Commission's Directorate General for International Cooperation and Development (DG DEVCO), the European Cooperation Agency is still generally referred to as Europe-Aid, so this study will use this name as it is the best known.
- ² Nor is the weight given to the distribution of funds by country, which is so often the case in community policy, made explicit.
- ³ Available in U.S. Government Publishing Office [GPO] (s.f.-b).
- ⁴ The model is in ADS Reference 303mav Certifications, Assurances, Other Statements of the Recipient and Solicitation Standard Provisions, partial revision 22 May 2017 (USAID 2017).
- ⁵ See as an example the concept of temporality rate used by the Spanish Ministry of Labour in "Servicio Público de Empleo Estatal" [SEPE], 2018.

References

- AECID. 2017. Proceso de Calificación de Organizaciones no Gubernamentales de Desarrollo, Manual de autoevaluación para la acreditación como ONGD calificada. Available online: <http://www.aecid.es/Centro-Documentacion/Documentos/documentos%20adjuntos/170810%201%20Manual%20Calif%20Nuevas.pdf> (accessed on 11 January 2019).
- AECID. 2019. Resolución de la Presidencia de la Agencia Española de Cooperación Internacional para el Desarrollo por la que se aprueba la convocatoria pública de concesión de subvenciones para la realización de proyectos de cooperación para el desarrollo, incluidos los de Educación para el Desarrollo en España, correspondiente al año 2019. Available online: <https://www.aecid.gob.es/es/Paginas/DetalleProcedimiento.aspx?idp=310> (accessed on 15 December 2019).
- Bach-Mortensen, Anders Malthe, and Paul Montgomery. 2018. What are the barriers and facilitators for third sector organisations (non-profits) to evaluate their services? A systematic review. *Systematic Review* 7: 2–15. [CrossRef] [PubMed]
- Berber, Philip, Patrick L. Brockett, William W. Cooper, Linda Golden, and Barnett R. Parker. 2011. Efficiency in fundraising and distributions to cause-related social profit enterprises. *Socio-Economic Planning Sciences* 45: 1–9. [CrossRef]
- Bisbe, Josep, and Joan Barrubés. 2012. El Cuadro de Mando Integral como instrumento para la evaluación y el seguimiento de la estrategia en las organizaciones sanitarias. *Revista Española de Cardiología* 65: 919–27. [CrossRef] [PubMed]
- BOE. 1998. Ley 23/1998, de 7 de julio, de Cooperación Internacional para el Desarrollo. Boletín Oficial del Estado (BOE) núm. 162 de 8 de julio de 1998. Beijing: BOE.
- BOE. 2011. Orden AEC/2909/2011, de 21 de octubre, por la que se establecen las bases para la concesión de subvenciones de cooperación internacional para el desarrollo. Boletín Oficial del Estado (BOE) núm.261, de 29 de octubre de 2011. Beijing: BOE.
- BOE. 2015. Real Decreto 193/2015 de 23 de marzo, por el que se aprueba el Reglamento del Registro de Organizaciones No Gubernamentales de Desarrollo. Boletín Oficial del Estado (BOE) núm.83, de 7 de abril de 2015. Beijing: BOE.
- Comisión Europea. 2016. Contratación pública y subvenciones para las acciones exteriores de la Unión Europea. Una guía práctica (Versión 2016.0-5 de enero de 2016) (PRAG). Available online: <http://ec.europa.eu/europeaid/prag> (accessed on 13 February 2018).
- Coordinadora de ONG para el Desarrollo-España. 2017. Informe de la Coordinadora sobre el sector de las ONGD. Edición 2017. Available online: <http://informe2017.coordinadoraongd.org/> (accessed on 19 April 2020).
- Cordery, Carolyn, and Rowena Sinclair. 2013. Measuring performance in the third sector. *Qualitative Research in Accounting & Management* 10: 196–212. [CrossRef]
- Europe-Aid Online Registration Service. 2008. Available online: <http://www.anong.org.uy/docs/noticias/docs%20convocatoria%20com%20europea/PADOR.pdf> (accessed on 4 July 2020).
- Ferreira, Marisa R., Amelia Carvalho, and Filipa Teixeira. 2017. Non-Governmental Development Organizations (NGDO) Performance and Funds—A Case Study. *Journal of Human Values* 23: 178–92. [CrossRef]
- Forbes, Daniel P. 1998. Measuring the unmeasurable: Empirical studies of nonprofit organization effectiveness from 1977 to 1997. *Nonprofit and Voluntary Sector Quarterly* 27: 183–202. [CrossRef]
- García Cebrián, Lucía Isabel, and Carmen Marcuello Cerbós. 2007. Eficiencia y captación de fondos en las organizaciones no gubernamentales para el desarrollo. *Revista de Economía Pública, Social y Cooperativa* 58: 221–49. Available online: <https://dialnet.unirioja.es/servlet/articulo?codigo=2470635> (accessed on 11 July 2020).
- Gálvez-Rodríguez, María del Mar, Carmen Caba Pérez, and Manuel López Godoy. 2016. NGOs efficiency and transparency policy: The Colombian case. *Innovar: Revista de Ciencias Administrativas y Sociales* 26: 67–82. [CrossRef]
- González Quintana, María José, and Encarnación Cañadas Molina. 2008. Los Indicadores de gestión y el cuadro de mando en las entidades no lucrativas. *CIRIEC-España, Revista de Economía Pública, Social y Cooperativa* 63: 227–52. Available online: <https://www.redalyc.org/pdf/174/17412307009.pdf> (accessed on 11 July 2020).
- Greiling, Dorothea. 2010. Balanced scorecard implementation in German non-profit organisations. *International Journal of Productivity and Performance Management* 59: 534–54. [CrossRef]
- Hernangómez Barahona, Juan, Natalia Martín Cruz, and Victor Martín Pérez. 2006. La relevancia del objetivo en la medida de la eficiencia. Un análisis para las ONGD españolas desde la teoría del comportamiento. *Boletín Económico del ICE* 2884: 55–68.
- Ishizaka, Alessio, Giuliano Resce, and Bertrand Mareschal. 2017. Visual management of performance with PROMETHEE productivity analysis. *Soft Computing* 22: 7325–38. [CrossRef]

- Kaplan, Robert S. 2001. Strategic performance and measurement in nonprofit organisations. *Nonprofit Management and Leadership* 11: 353–70. Available online: <http://docshare02.docshare.tips/files/13452/134524942.pdf> (accessed on 11 July 2020).
- Lecy, Jesse D., Hans Peter Schmitz, and Haley Swedlund. 2012. Non-governmental and not-for-profit organizational effectiveness: A modern synthesis. *Voluntas: International Journal of Voluntary and Nonprofit Organizations* 23: 434–57. [CrossRef]
- Pérez Martín, Victor, Natalia Martín Cruz, and Cesar Gámez Alcalde. 2012. La eficiencia y el misreporting contable en las ONGD españolas. Análisis de proyectos de cooperación internacional para el desarrollo. *Academia: Revista Latinoamericana de Administración* 51: 1–14. Available online: <https://www.redalyc.org/pdf/716/71625040002.pdf> (accessed on 12 March 2020).
- Martín Pérez, Victor, Natalia Martín Cruz, and Maria Jose Serrano Rodríguez. 2015. Eficiencia de las ayudas al desarrollo: El caso español. *Revista Venezolana de Gerencia (RVG)* 20: 285–311. [CrossRef]
- Medina Rey, Jose Maria. 2009. Transparencia y buen gobierno en las ONGD 'S. *Revista Española del Tercer Sector* 11: 93–113. Available online: <http://www.plataformatercersector.es/sites/default/files/N11%20RETS%20Las%20ONG%20de%20Desarrollo.pdf> (accessed on 18 July 2020).
- Mouchamps, Hugues. 2014. Weighing elephants with kitchen scales. *International Journal of Productivity and Performance Management* 63: 727–45. [CrossRef]
- Moxham, Claire. 2014. Understanding third sector performance measurement system design: A literature review. *International Journal of Productivity and Performance Management* 63: 704–26. [CrossRef]
- Registro federal de contratistas de EE UU (USFCR). 2020. Start a SAM Registration. Available online: https://usfcr.com/sam-registration/?utm_medium=ppc&utm_term=system%20for%20award%20management&utm_campaign=SAM&utm_source=adwords&hsa_tgt=kwd-37660679064&hsa_ad=442167707643&hsa_src=g&hsa_mt=b&hsa_ver=3&hsa_cam=122161541&hsa_acc=1449812725&hsa_kw=system%20for%20award%20management&hsa_grp=5341822301&hsa_net=adwords&gclid=Cj0KCQjwZTS5BRcdARIsAGEX0zlhUxrGhPn1Umpcxd70Z6hD850pE8upyfXwyIIBFX3piF13pe20fosaAtIIEALw_wcB (accessed on 13 July 2020).
- Retolaza, Jose Luis. 2010. Factores Estratégicos de Éxito en las Empresas de Inserción. Ph.D. thesis, Universidad del País Vasco, UPV/EHU, Leioa, Spain.
- Ribar, David C., and Mark O. Wilhelm. 2002. Altruistic and joy-of-giving motivations in charitable behavior. *Journal of Political Economy* 110: 425–57. Available online: <https://econweb.ucsd.edu/~jandreon/PhilanthropyAndFundraising/Volume%201/16%20Ribar%20Wilhelm%202002.pdf> (accessed on 2 June 2020).
- Sainz, Hector. 2007. Venturas y desventuras del enfoque del marco lógico. *Revista Española de Desarrollo y Cooperación* 20: 133–50.
- USAID. 2012. Non-U.S. Organization Pre-Award Survey, Guidelines and Support, Additional Help for ADS Chapter 303. Available online: <https://www.usaid.gov/sites/default/files/documents/1868/303sam.pdf> (accessed on 23 March 2018).
- USAID. 2017. ADS Reference 303maa, Standard Provisions for U.S. Nongovernmental Organizations, partial revisión 22 de mayo de 2017. Available online: <https://www.usaid.gov/ads/policy/300/303maa> (accessed on 8 March 2018).
- USAID. 2018. Municipal Waste Recycling Program (MWRP) to Reduce Plastics Pollution of the Oceans–Indonesia, Philippines, Sri Lanka, and Vietnam. Available online: https://www.usaid.gov/sites/default/files/documents/1861/MWRP_APS_as_Amended_2018_01.pdf (accessed on 1 August 2020).
- Non-Us Organization Pre-Award Survey (NUPAS). n.d. Available online: <https://www.usaid.gov/sites/default/files/documents/1868/303sama1.docx> (accessed on 12 March 2018).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

One Man's Bubble Is Another Man's Rational Behavior: Comparing Alternative Macroeconomic Hypotheses for the US Housing Market

Anastasios G. Malliaris ^{1,*}, Mary Malliaris ¹ and Mark S. Rzepczynski ²

¹ Quinlan School of Business, Loyola University Chicago, 16 E. Pearson, Chicago, IL 60611, USA; mmallia@luc.edu

² AMPHI Research and Trading, LLC, Concord, MA 01742, USA; mrzepczynski@amphitrading.com

* Correspondence: tmallia@luc.edu

Abstract: Competing macroeconomic hypotheses have been developed to explain the US housing market and possible bubble behavior. We employ both seasonally adjusted (SA) and non-seasonally adjusted (NSA) monthly data for about 30 independent variables to examine alternative macro hypotheses for home prices. Using a neural network model as an atheoretical non-linear approach to capture the relative importance of alternative macro variables, we show that these hypotheses generate different macro relevance. As an alternative to testing housing time series, we focus on bubble identification being hypothesis dependent. Model forecast errors (residuals) identify the potential presence of bubbles through standardized residual CUSUM tests for structural breaks. By testing for housing bubbles from these unstructured models, we generate conclusions on the presence of bubbles prior to the Great Financial Crisis and the post-pandemic periods. Competing macro hypotheses or narratives will generate different conclusions on the presence of bubbles and create bubble identification issues.

Keywords: housing prices; Great Financial Crisis; quantitative easing; housing bubbles; neural network methodology

JEL Classification: E10; E32; E44; E58; E62; R31

Citation: Malliaris, Anastasios G., Mary Malliaris, and Mark S. Rzepczynski. 2024. One Man's Bubble Is Another Man's Rational Behavior: Comparing Alternative Macroeconomic Hypotheses for the US Housing Market. *Journal of Risk and Financial Management* 17: 349. <https://doi.org/10.3390/jrfm17080349>

Academic Editor: Susan Wachter

Received: 4 July 2024

Revised: 4 August 2024

Accepted: 8 August 2024

Published: 12 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Housing dynamics has become a subject of intense interest since the Global Financial Crisis (GFC) via the potential for “bubbles” that may create macro financial instability. However, there is limited agreement on when housing bubbles occur, especially when framed by macro drivers or hypotheses. Any identification of a bubble based on macro drivers is dependent on the narrative and variables employed. Bubble identification is a joint test of the model used to explain housing price dynamics and the actual presence of a bubble. Two different models may draw different conclusions on the presence of a bubble.

Even with the knowledge gained since the GFC, we have seen significant housing price increases as measured by the S&P CoreLogic Case–Shiller U.S. National Home Price Index, which may represent another housing bubble. After the GFC, housing prices increased from February 2011 to March 2020 by over 55 percent during a period of relatively slow economic growth and near-zero short rates. The dramatic increases in liquidity by the Fed during and after March 2020 and the easy fiscal policies employed to counteract the COVID-19 recession offered further stimulus to housing. The housing index rapidly climbed by over 40% in just over 2 years. Housing prices have reached new highs with economic growth not significantly higher than the pre-GFC period even with new post-GFC macro-prudential policies to reduce speculative housing excess. A key housing question is whether this is a rational response to macro factors or an irrational bubble.

While significant research has been conducted on identifying housing bubbles and price extremes, these macro finance works have focused on time series (discounted present value) properties and have neglected the macro drivers of housing. We employ a different two-step approach by first testing a set of alternative narratives or theories concerning housing price dynamics using a non-linear atheoretical neural network model to compare narratives that can explain housing prices, and second, identifying housing bubbles based on these hypotheses.

From a set of competing macro hypotheses proposed to explain housing dynamics, we test for bubbles using a CUSUM test for structural breaks and conclude that the presence of asset bubbles is macro narrative dependent. While there is general agreement on a housing bubble preceding the GFC, our analysis extends through 2022 to determine whether prominent explanations from the GFC cycle are also applicable to the housing cycle during COVID-19.

We are not trying to definitively identify the presence of a bubble, but given that certain assets, such as housing, realize price booms and busts, can different competing macro hypotheses provide useful insights? If the residual CUSUM reaches an extreme, there is evidence that a bubble may have existed, conditional on the model employed. Conversely, if a model does not have significant positive CUSUM residuals, there is evidence that specific macro factors can explain the large price moves and there is no bubble.

From training and testing features across extended housing data, we draw inferences on: (1) the quality of hypotheses to describe housing prices, and (2) the potential presence of bubbles based on the structure of forecast errors. Our methodology of combining machine learning with tools for measuring structural breaks to identify bubbles can serve as an operational framework for problems with complex alternative hypotheses.

After a literature review on housing price dynamics during the GFC and the measurement of housing bubbles in Section 2, we propose a neural network (NN) methodology in Section 3 to investigate alternative hypotheses formulated in Section 4. These hypotheses and related analysis associated with several economic variables that impact house prices provide comparisons of conflicting views that promote different narratives used to explain housing price cycles. The main conclusions are summarized in the last section and focus on the presence of bubbles through competing macro housing hypotheses.

2. Literature Review

Duca et al. (2021) present a comprehensive literature survey of what drives house prices with an emphasis on papers written since 2007. Housing prices are determined by complex relationships with macroeconomic variables, financial and credit conditions, monetary policies, globalization, microeconomic supply and demand factors, regulation, and asset price bubbles, all playing roles. This work highlights the problem of competing hypotheses trying to explain housing dynamics.

Leamer (2015) establishes that housing belongs to the group of macroeconomic business cycle variables (GDP, consumption, investment, industrial productions, disposable income, and unemployment) linked to the business cycle. He offers evidence that from 1950 to 2000, 9 out of 11 recessions were preceded by housing price declines.

The impact of monetary policy on housing prices has been proposed by Taylor (2007, 2009, 2010), who argues that an easy monetary policy fueled the housing bubble. McDonald and Stokes (2013) also discuss the impact of monetary policy on housing, and Bhar and Malliaris (2021) compare the role of monetary policy on housing both during the GFC and COVID-19. Corsi and Sornette (2014) argue that asset bubbles usually occur during periods of excess liquidity. Bernanke (2010) disputes the thesis that the housing bubble during 2000–2006 was driven by an easy monetary policy. This easy money issue has not been addressed after 2010.

Bernanke (2005, 2007a, 2007b, 2009) proposes that the housing bubble was fueled by lower longer-term interest rates for 10-Year Treasury Notes and 30-Year Treasury Bonds. Such low longer-term interest rates also cause housing mortgage rates to decline. These

lower rates were associated with a global saving glut. Sa and Wieladex (2015) and Steinberg (2019) investigate empirically the role of the global saving glut, and Evgenidis and Malliaris (2023) show that the global saving glut contributed to the housing bubble prior to the GFC but did not drive the rapid increases in housing prices during and after COVID-19. Beyond the impact of monetary policy and the global saving glut on mortgage rates, the housing micro bubble literature focuses not on whether prices rose too much but rather on why and what role borrower price expectations versus lender easing of underwriting played in enabling and separately contributing to excess prices. This micro approach is developed in Levitin et al. (2020), who show that as the volume of private-label mortgage-backed securitized loans increased and lending terms eased, risk premiums failed to price the increase in risk. This easy credit supply combined with bullish house price expectations contributed to the pre-GFC housing bubble and is a compelling argument. Also, the easy credit supply narrative on a micro level is consistent with the global saving glut macro argument.

Little is available in the literature comparing the housing bubble pre-GFC and post. Evgenidis and Malliaris (2023) study the role of monetary policy on housing price extremes, pre-GFC and during COVID-19, and find different drivers for each. Bhar et al. (2024) offer comparative analysis of housing prices pre-GFC and during COVID-19 by focusing on the stability of coefficients and its impact on housing forecasts without focusing on housing bubbles. Specifically, Bhar et al. (2024) propose 5 hypotheses to explain house prices using 15 economic variables and employ standard dynamic econometric techniques to test for significance of independent variables across regime changes. The general conclusion that emerges is that both SA and NSA house prices over a long period from 1987 to the end of COVID-19 were driven by the nonlinearity of coefficients of 15 economic variables, with varying impact across certain sub-periods.

The present paper extends Bhar et al. (2024) first by proposing 10 hypotheses entailing 30 independent variables to contribute towards a more complete understanding of house prices; second, by employing an NN methodology to provide an alternative approach not constrained by linear estimation; and third, by evaluating models to assess any evidence of the presence of bubbles. Instead of testing a formal structural linear model or a simple vector autocorrelated model, an NN framework is employed to find those variables that have the greatest impact on housing variability and the macro themes that have the highest predictive value. Using this common modeling methodology and sample periods allows us to make comparisons among key independent variables driving housing prices and draw conclusions on different strands of macro housing research.

The structure of forecast errors from our housing themes is used to test for bubbles. Any bubble discussion is dependent on the model used to explain price dynamics. Model errors trained on our data sets should have an expected value of zero and be iid. Systematic deviations of standardized residuals can identify potential model failure, structural breaks, and insights into the possibility of bubbles. The behavior of the residual CUSUM along with formal testing allows us to identify structured deviations from competing models as potential bubbles. Across different model themes, we find evidence for structural breaks surrounding the GFC, which suggests that these models or themes inadequately describe housing cycle dynamics. Those who use specific themes to describe housing market dynamics may inappropriately identify a housing bubble that may not exist based on alternative theses.

A bubble has been simply described by Kindleberger (1996, p. 13) as “an upward price movement over an extended range that then implodes”. Similar definitions emphasize unsustainable excessive price increases. A bubble can also be defined as an asset price that exceeds an asset’s fundamental value because current owners believe they can resell the asset at an even higher future price. A bubble cannot be separated from a model or valuation or from the drivers that impact valuation.

Books such as Kindleberger (1996) and Evanoff et al. (2012), and survey papers such as Scherbina and Schlusche (2014) give a detailed analysis of financial modeling of asset

bubbles. These models provide alternative hypotheses based on asymmetric information between traders, the interaction between rational and behavioral traders, limits to arbitrage, investors holding heterogeneous beliefs, and disagreements about fundamental values.

Bubble research has often focused on finding extreme price movements by measuring forms of exponential growth. See, for example, Sornette and Cauwels (2014) for a review of the price-based approach to bubble identification. This strain of research uses techniques from operations research; see Sornette (2003), Corsi and Sornette (2014), Sornette et al. (2017), and Ziemba et al. (2017). Excess returns over compact timeframes, the focus of the exponential work, are less applicable when national housing price data with differences in quality are calculated monthly.

The study of housing price bubbles has been extensive with special emphasis on Phillips-type right-tailed unit root tests, such as the Supremum Augmented Dickey Fuller (SADF) and Generalized SADF (GSADF). Hu (2023) offers a review of this approach. The work of Phillips and Shi (2018), Martinez-Garcia and Grossman (2018), and Shi and Phillips (2020) has led to real-time monitoring of housing bubbles around the world by the Dallas Federal Reserve Bank.¹ However, these exuberance or bubble indicators focus on real house prices and price-to-rent fundamentals and do not address the question of whether macro fundamental hypotheses can be used to explain the price changes or exuberance.

Otero et al. (2022) undertake Monte Carlo simulation experiments to examine the effect of changing the frequency of observations and the data span and find that when a series is characterized by multiple bubbles, periodically collapsing, decreasing the frequency of observations is associated with profound power losses for the test. This motivates us to use a long series of data with the empirical observation of one collapsing housing bubble after the Global Financial Crisis. Holmes et al. (2011) investigate regional house price convergence in the United States, complementing our national house price bubble behavior.

West (1988) suggests that fads may contribute to bubbles, but in our analysis such fads do not play a critical role. The theoretical modeling of Kiselev and Ryzhik (2010) incorporates three main factors as components for asset price bubbles. These components are the tendency of mean reversion to a stable value, speculative social response triggered by trend following, and random fluctuations. The interplay of these three forces may lead to bubble formation and collapse. Tarlie et al. (2022) discuss explosive behavior of stocks relative to valuation. We do not test or look at a valuation model for house bubbles but rather focus on macro drivers as explanations of price change. LeRoy (2012) reminds us of the role of holding periods, with housing generally being a long hold.

Our premise is that any bubble discussion should be a two-step process of first identifying and testing alternative feature hypotheses as explanations of price extremes, and then testing whether price breaks from these explanatory features serve as a bubble identifier. Albeit more difficult to form definitive conclusions, this approach allows commentary on the value of alternative macro narratives.

3. A Neural Network Framework

A supervised learning approach using artificial neural networks as an alternative to linear regression is employed to rank the importance of a set of macro and market variables that may have complex non-linear relationships with housing prices.² Instead of forming a single structural model, we measure the relative importance of variables associated with specific narratives for the housing market. By focusing on the relative importance of variables and the ability of specific models to explain housing prices, we address core macro housing issues. Our atheoretical restriction-free approach compares specific hypotheses and measures relative validity and the importance of macro housing themes through differences in model prediction errors.³

The supervised neural network (NN) algorithm feeds data forward through layers of nodes or neurons. Neural networks adapt based on errors, and then improve by making small adjustments to weights. It is a repetitive process that continues until the error meets some pre-specified minimum. Our network is composed of three layers: nodes for each

input, an intermediary layer, which is the “hidden” layer that contains nodes connected to both the input and output nodes, and a third layer consisting of a node for the output value. Each input node is connected to every node in the hidden layer and is multiplied by a weight. Each hidden layer node also has a weighted connection to the output node. Initial weights are assigned randomly and adjusted to minimize prediction error. The weighted connections that arrive at a node are summed and sent through a hyperbolic tangent function with the output from each hidden layer node serving as an input for the next layer. The final node sums the weighted inputs from the hidden layer nodes and applies a hyperbolic tangent function to this sum where this last layer node becomes the model’s target value prediction. An illustration of this NN methodology is presented in Malliaris and Malliaris (2021). SPSS Modeler automatically computes the optimal size of the hidden layer by trying various sizes and selecting the one that results in the minimum test error. If multiple networks have the same error, the one with the fewest hidden nodes is retained. Since the objective in these hypotheses was to best model the variable relationships, all rows for each hypothesis were used in building that model. The specific architecture used in the network for each hypothesis varies depending on the input data. At the end of each figure description, the network architecture for that model is specified. For example, Figure 1 specifies: Network Architecture: 7 input nodes, 2 hidden layer nodes, 1 output node.

The model is trained through an error backpropagation algorithm to compute the first partial derivatives of the error function with respect to the weights. Weights are adjusted based on the error, using the gradient descent method and the generalized delta rule. A change to the weight calculated by the generalized delta rule would be:

$$w_{ji}^{(l)}(n+1) = w_{ji}^{(l)}(n) + \alpha [w_{ji}^{(l)}(n) - w_{ji}^{(l)}(n-1)] + \eta \delta_j^{(l)}(n) y_i^{(l-1)}(n)$$

where $w_{ji}^{(l)}$ indicates weight w , in layer l , in neuron j , fed from neuron i in the previous layer. The iteration is represented by n , with a learning-rate parameter η and momentum of α . See Haykin (1994, p. 156) for a detailed discussion of this process. This procedure is repeated until either the total error is below a specified level or the number of data passes has exceeded a specified maximum.

A sensitivity analysis is conducted after training to determine the influence of each feature variable on the output (housing prices) through varying the input variable values while holding the other variables fixed. By observing the value of the mean squared error corresponding to each agitated input, the model can determine the output sensitivity to that input. The sensitivity analysis computes the decrease in variance of the target variable that can be attributed to each input variable. Saltelli et al. (2004) show that this normalized sensitivity is a good measure to rank the predictors in order of importance for any combination of interaction and non-orthogonality among predictors. The input variables are ranked by the effect of their changes on the output variable, which generates a predictor importance ranking. The sensitivity measure ranking the predictors is $S_i = V_i/V(Y) = V(E(Y|X_i))/V(Y)$ where $V(Y)$ represents the output variance and X represents inputs. The predictor importance is the normalized sensitivity $VI_i = S_i/\sum S_j$.

Methodology for Testing Bubbles

While significant work has focused on price data time series, the unit root problem, a bubble cannot be divorced from a valuation model. A perceived bubble may not be a bubble if we find exogenous variables that explain price behavior without any residual bias; however, a bubble may be identified by the deviations (prediction errors) away from a model, a structural break from model behavior. Homm and Breitung (2012) associated the structure of forecast errors as an identification of bubbles; however, most of this work has focused on price series and not specific models using exogenous variables. The pattern of time series residuals, deviations from a model’s forecast, can be used as a bubble test. Hence, CUSUM tests can measure specific structural breaks in a model. Bubble identification is

conditional on the model used. Ploberger and Krämer (1992) test OLS residuals to identify and locate breaks.

We explicitly link an atheoretical machine learning framework across different sets of features which represent different macro narratives to test the presence of housing market bubbles conditional on the model selected through tests of residuals. Inspection of the CUSUM time series from our atheoretical NN models, along with tests of critical values, can identify whether model failure is consistent with when bubbles are said to have existed. Cumulative deviations from a model over the training set is a strong test of bubbles because the model has all information prior to and after the bubble. This is not a test of a model's out-of-sample forecast ability but a focus on whether a set of variables can adequately find a fit that will generate forecast errors that are iid with an expected value of zero. A model can have high forecast errors but have an expected mean of zero; however, a CUSUM, recursive residual test can identify structural errors that can be viewed as bubbles. CUSUMQ test results are also available upon request.

4. Hypotheses and Analysis

We investigate the influence of macroeconomic variables on monthly housing prices from January 1987 to June 2022. The independent and macro variables are organized into two categories: variables that are Not Seasonally Adjusted (NSA) and variables that are Seasonally Adjusted (SA). There are both economic and statistical reasons for this distinction. All housing transactions occur with unadjusted prices. Anticipations and expectations are also not seasonally adjusted by builders, sellers, and buyers. Statistically, mixing NSA and SA impacts estimations since the SA data incorporates a filter with its own dynamics. By offering two sets of results, we enrich our understanding of the dynamics played by different independent variables. When important variables are available in only one mode, only NSA or only SA, we allow for some mixing with some caution in interpreting results. The target Case–Shiller home price variable is available as either NSA or SA.

Table 1 provides an overview for all our 10 major hypotheses and sets of variables. We generate the ranked feature importance within each hypothesis since all model variables are scaled between 0 and 1, as well as statistics on model errors. Hence, the values measure the amount of volatility explained by a given feature.

Table 1. Hypotheses and variables (features) for neural network tests. All data from FRED.

Hypotheses	Variables
Housing and Macroeconomic Variables	
H1 NSA Macroeconomic variables	Consumer Price Index All (CPI) annualized change, Consumer Price Index Rent annualized change (CPIR), Non-Farm payroll number (NFP), Industrial Production annualized change (IP), 30-year Mortgage Rate (MORT), macro news (MACRO), and Economic Policy Uncertainty Index (EPU)
H2 SA Macroeconomic variables	Durable goods (DGOOD), Disposable income (DINC), Trade balance (TBAL), Employment/population ratio (EMRATIO), Unemployment level (UNEMPLOY), Industrial production (IP)
Housing driven by monetary policy	
H3 NSA credit variables	30-year Mortgage Rate (MORT), Treasury 10-year/2-year spread (TSPREAD), Fed Funds (FFUND), EPU monetary index (EPUM)
H4 NSA credit with Fed assets	30-year Mortgage Rate (MORT), Treasury 10-year/2-year spread (TSPREAD), Fed Funds (FFUND), EPU monetary index (EPUM), FedAssets (FEDA)
Housing and global savings glut	
H5 NSA target variable	30-year Mortgage Rate (MORT), Trade balance (TBAL), Fed Funds (FFUND), FedAssets (FEDA)

Table 1. Cont.

Hypotheses	Variables
H6 SA target variable	30-year Mortgage Rate (MORT), Trade balance (TBAL), Fed Funds (FFUND), FedAssets (FEDA)
Housing supply and demand	
H7 NSA variables	Monthly supply of new houses (HSUPPLY), new single-family houses sold (HSOLD), single-family housing units completed (HCOMPL), industrial production (IP), non-farm payroll (NFP)
H8 SA variables	Monthly supply of new houses (HSUPPLY), new single-family houses sold (HSOLD), single-family housing units completed (HCOMPL), disposable income (DINC), durable goods (DGOOD), trade balance (TBAL)
H9 Extends H7 by adding expectations	Monthly supply of new houses (HSUPPLY), new single-family houses sold (HSOLD), single-family housing units completed (HCOMPL), industrial production (IP), non-farm payroll (NFP), KC Fed stress index (STRESS), University of Michigan consumer sentiment (MSCENT), single-family units started/population level (HPOP), housing authorized but not started (HAUTH)
Housing and price extrapolation	
H10 Price lags and momentum	Case–Shiller prices lagged 1 period (CSLAG1), Case–Shiller prices lagged 6 period (CSLAG6), Fed funds (FFUND), housing units authorized but not started (HAUTH), single-family housing units completed (HCOMPL), Chicago Fed financial conditions (CFINCON), University of Michigan consumer sentiment (MCSENT), Michigan inflation expectations (MINFEX)

Comparing the NSA and SA analysis, cyclical housing behavior is driven by traditional macroeconomic variables such as industrial production, durable goods, disposable income, non-farm payroll, and two special factors that are relevant to housing: CPI rent and trade balance. For NSA data, the three most important cyclical variables are the CPI rent, industrial production, and non-farm payroll while for SA data, durable goods, disposable income, and trade balance emerge as most important. The SA model has half the mean absolute error and half the standard deviation as the NSA test. Contemporaneous business cycle effects include a trade component.

These two hypotheses conclude that national US economic cycles play an important part in housing cycles. Both NSA and SA data confirm that different cyclical macroeconomic variables are connected to housing prices. Housing cycles, as expected, are congruent with the business cycle.

Hypothesis 1. *Case–Shiller National Home Price Index NSA is driven by macro variables. The non-seasonally adjusted (NSA) data set ranges from January 1987 through June 2022 and includes the consumer price all index NSA (CPI), consumer price rent index NSA (CPIR), Non-farm payroll NSA (NFP), industrial production NSA (IP), 30-year Mortgage Rate (MORT), macro news (MACRO), and the Economic Policy Uncertainty Index (EPU) to capture overall macro risk. These independent variables are chosen among a large set of macro data available to capture relationships between house prices and consumer prices, the consumer price index measuring rent as a relevant alternative opportunity cost to purchase, non-farm payroll, and industrial production are proxies for GDP. The 30-year mortgage rate is a primary variable that measures house financing and affordability. Macro news and the economic policy uncertainty index serve as proxies for the scope of financial conditions.*

The results in Figure 1 generate a notable surprise with the most important variable, the CPI rent component (weight 0.341), which proxies for the cost of overall housing but does not play a central role in business cycle aggregate demand analysis. Two important features, employment and industrial production, key business cycle variables, only account for approximately 40% of the housing price variability.

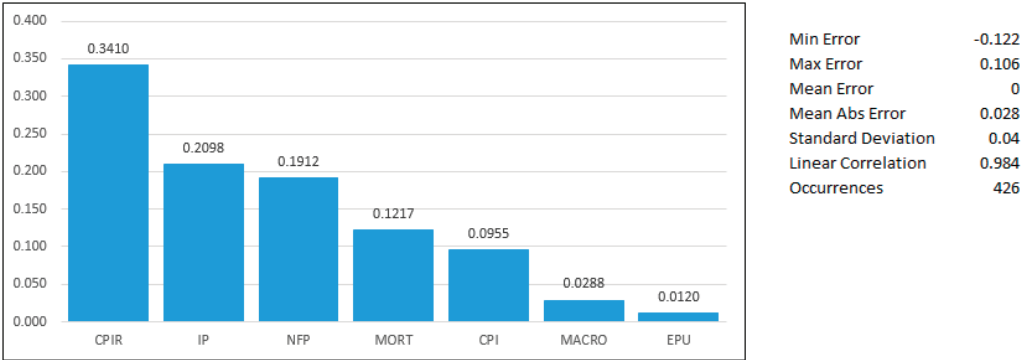


Figure 1. H1. Case-Shiller National Home Price Index NSA and macro variables sensitivity. Period: Jan 1987-June 2022. Source FRED database: Consumer price index all (CPI) annualized change, Consumer price index rent annualized change (CPIR), Non-Farm payroll number (NFP), Industrial Production annualized change (IP), 30-year mortgage rate (MORT), macro news (MACRO), and Economic policy uncertainty Index (EPU). Data scaled between 0 and 1 and sensitivities sum to 1. Network Architecture: 7 input nodes, 2 hidden layer nodes, 1 output node.

Hypothesis 2. Case-Shiller National Home Price Index SA is driven by different macro variables. Seasonally adjusted macroeconomic variables are introduced in this hypothesis for the period 1992-June 2022, and the dependent variable is also seasonally adjusted. Disposable income (DINC), durable goods (DGOOD), industrial production (IP), trade balance (TBAL), and unemployment (UNEMPLOY) are all proxies for the business cycles; employment to population ratio (EMRATIO) is introduced to adjust employment relative to population size. Durable goods, disposable income, and trade balance are the three most important drivers.

Results are presented in Figure 2.

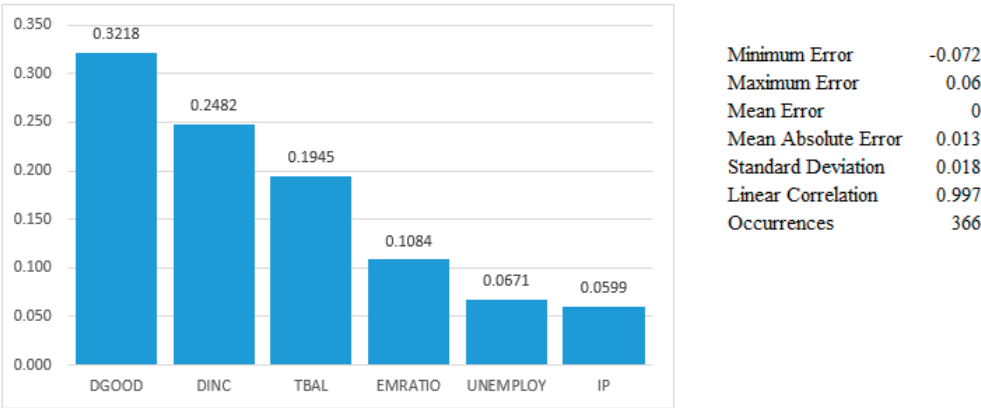


Figure 2. H2. Case-Shiller National Home Price Index SA and other macro variable sensitivity. Period: Jan 1992-June 2022. Source FRED database: Durable goods (DGOOD), Disposable income (DINC), Trade balance (TBAL), Employment/population ratio (EMRATIO), Unemployment level (UNEMPLOY), Industrial production (IP). Data scaled between 0 and 1 and sensitivities sum to 1. Network Architecture: 6 input nodes, 3 hidden layer nodes, 1 output node.

Hypotheses 3 and 4 address the impact of monetary policy on house prices. In Hypothesis 3, we test the role of Fed funds with other financial variables, while in Hypothesis 4,

we place an emphasis on Fed assets accumulated via QE. Sample periods differ since there is little movement in Fed’s balance sheet (no QE) prior to the GFC.

NSA housing prices for the entire sample are driven by the 30-year mortgage rate and the difference between the 10-year and 2-Year Treasury rate, which measure the slope of the yield curve, while the Fed funds rate ranks as the third factor. When a similar exercise is performed for a shorter period of 2003 to June 2022, Fed assets become most important followed by Fed funds. These two variables dominate the 30-year mortgage rate as the key drivers of housing behavior. Hypotheses 3 and 4 confirm that the housing market is impacted by monetary policy, initially and up to the GFC through Fed funds policy and after the GFC by QE.

Hypothesis 3. *Housing prices NSA and monetary policy. The Case-Shiller National Home Price Index NSA is impacted by monetary policy expressed by the Fed funds, the 10-year Treasury minus 2-year spread, 30-year mortgage rate, and an economic policy uncertainty index to capture the changes in policy expectations from 1987 to June 2022.*

These results in Figure 3 conclude that NSA house prices are driven by the 30-year mortgage rate, the term structure of interest rates represented by the 10-year less the 2-year spread, and Fed funds rate. However, monetary policy variables have much less explanatory power than the macroeconomic variables of Hypotheses 1 and 2, with greater dispersion in errors as Figures 1–3 illustrate.

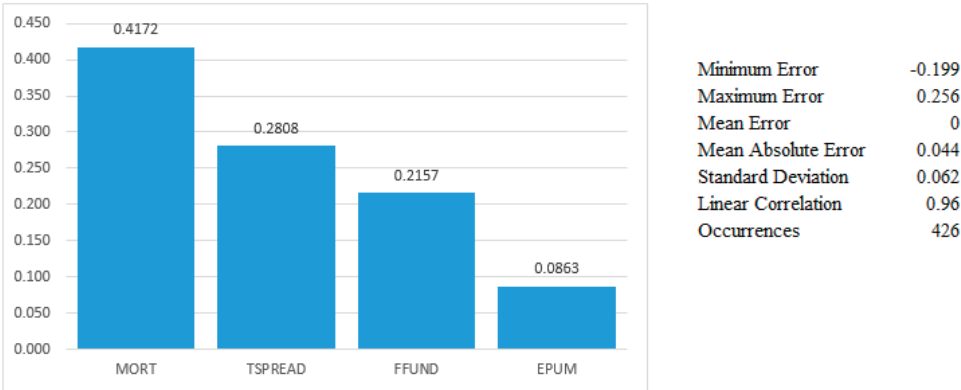


Figure 3. H3. Case-Shiller National Home Price Index NSA monetary policy sensitivity. Period: Jan 1987–June 2022. Source FRED database: 30-year mortgage (MORT), Treasury 10-year/2-year spread (TSPREAD), Fed Funds (FFUND), EPU monetary index (EPUM). Data scaled between 0 and 1 and sensitivities sum to 1. Network Architecture: 4 input nodes, 6 hidden layer nodes, 1 output node.

Hypothesis 4. *Housing prices and monetary policy including Fed assets. We study the behavior of the same dependent variable by introducing assets on the Fed balance sheet, which becomes the most important feature. The low mean absolute error for the post-GFC monetary model theme with Fed assets even with the Fed funds rate tied closely to zero is especially notable.*

The results from Hypothesis 4 are presented in Figure 4 that illustrate the important role of Fed assets.

Hypotheses 5 and 6 address the role of global saving glut. We compare drivers that are NSA as in Hypotheses 3 and 4 with the global savings glut using the trade balance as a proxy that is only available as SA. To moderate this statistical difficulty of mixing SA and NSA data, we use two models, expressed as H5 and H6.

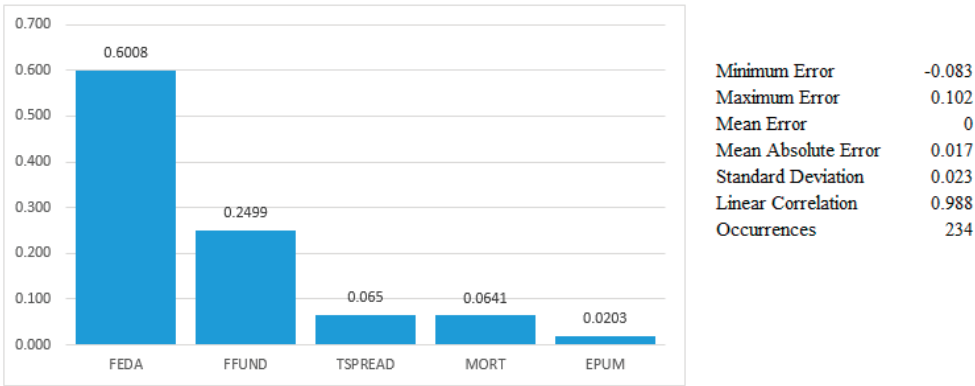


Figure 4. H4. Case-Shiller National Home Price Index NSA monetary policy sensitivity. Period: Jan 1992-June 2022. Source FRED database: 30-year mortgage (MORT), Treasury 10-year/2-year spread (TSPREAD), Fed Funds (FFUND), EPU monetary index (EPUM), Fed assets (FEDA). Data scaled between 0 and 1 and sensitivities sum to 1. Network Architecture: 5 input nodes, 2 hidden layer nodes, 1 output node.

Hypothesis 5. *Housing prices (NSA) with monetary policy and trade balance.*

Hypothesis 6. *Housing prices (SA) with monetary policy and trade balance.*

With H5, Case-Shiller NSA is explained in Figure 5 by Fed funds, Fed total assets, the 30-year mortgage rate (all three NSA), and the trade balance (SA). The second case, H6, considers the Case-Shiller SA dependent variable with the same independent variables to test whether the global saving glut (trade balance) emerges as an important variable. Adjusting for seasonality does not influence the results.

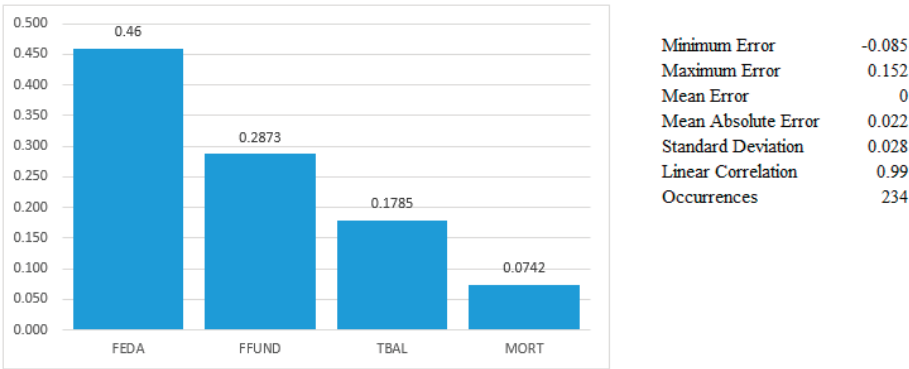


Figure 5. H5. Case-Shiller National Home Price Index NSA monetary policy sensitivity. Period: Jan 2003-June 2022. Source FRED database: 30-year mortgage (MORT), Trade Balance (TBAL), Fed Funds (FFUND), EPU monetary index (EPUM), Fed assets (FEDA). Data scaled between 0 and 1 and sensitivities sum to 1. Network Architecture: 4 input nodes, 4 hidden layer nodes, 1 output node.

Figures 5 and 6 show the importance of Fed assets and Fed Funds as drivers of home prices. We next examine the housing market with specific supply and demand information based on the recent housing research surveys of Glaeser and Gyourko (2018) and Duca et al. (2021). Erdmann (2023) focuses on supply and demand fundamentals in metropolitan areas. These results confirm the earlier analysis, where SA cyclical variables such as

disposable income, trade balances, and durable goods spending drive housing prices. Supply variables represented by houses started, houses completed, and houses sold play a lesser role. The supply of housing appears less elastic, and housing prices are driven by cyclical factors caused by demand shifts. Figures 7 and 8 show the impact of the variables proposed in hypotheses 7 and 8.

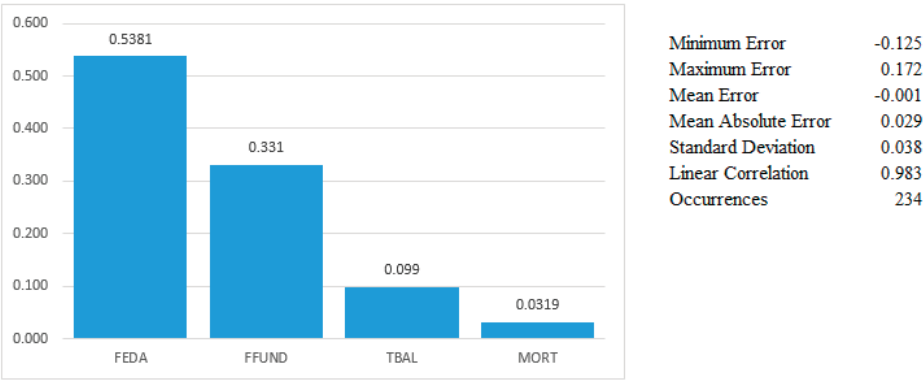


Figure 6. H6. Case-Shiller National Home Price Index SA monetary policy sensitivity. Period: Jan 2003-June 2022. Source FRED database: 30-year mortgage (MORT), Trade Balance (TBAL), Fed Funds (FFUND), Fed assets (FEDA). Data scaled between 0 and 1 and sensitivities sum to 1. Network Architecture: 4 input nodes, 3 hidden layer nodes, 1 output node.

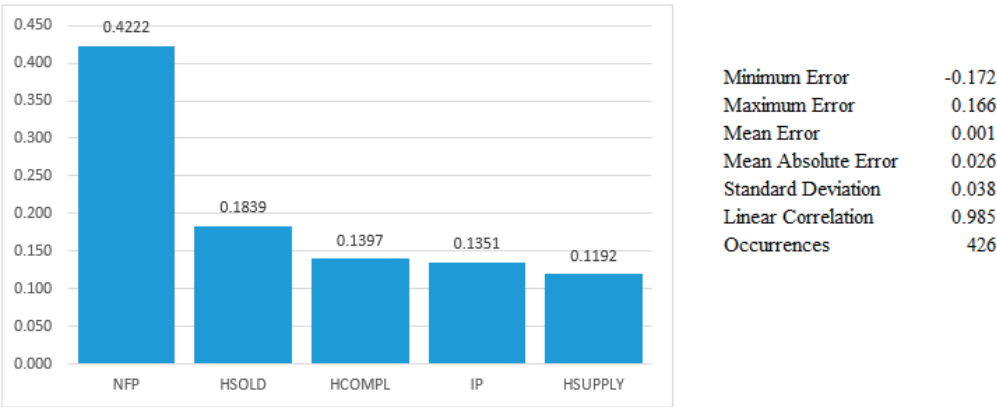


Figure 7. H7. Housing prices (NSA), micro supply and macro demand factors. Period: Jan 1987- June 2022. Source FRED database: Monthly supply of new houses (HSUPPLY), new single-family houses sold (HSOLD), single-family housing units completed (HCOMPL), industrial production (IP), and non-farm payroll (NFP). Variables are NSA. Data scaled between 0 and 1 and sensitivities sum to 1. Network Architecture: 5 input nodes, 4 hidden layer nodes, 1 output node.

Hypothesis 7. Housing prices (NSA), micro supply and macro demand factors. Case-Shiller NSA housing prices are explained by supply variables: monthly supply of new houses, new single-family houses sold, single-family housing units completed, as well as industrial production and non-farm payroll. There are no data for explicit housing demand other than proxies for aggregate demand which focus on the business cycle. Results using NSA variables suggest that housing fundamentals such as houses completed, houses supplied, and houses sold are important influences on prices, but the demand side through employment and, to a lesser extent, industrial production are also key drivers.

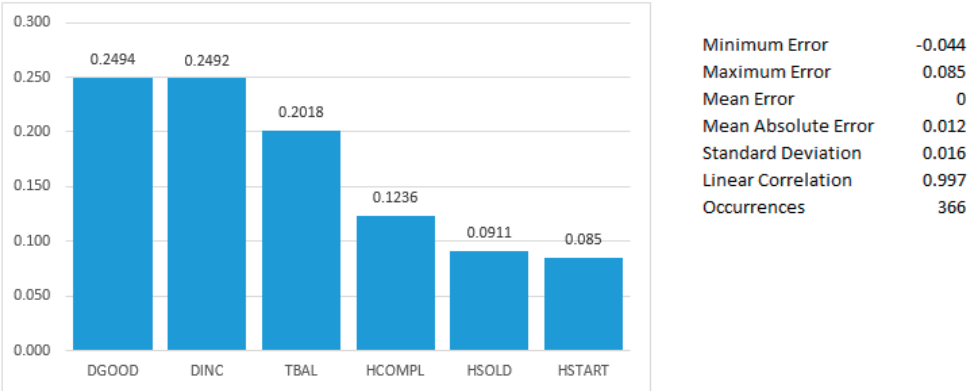


Figure 8. H8. Housing Prices (SA) and micro supply and demand factors. Period: Jan 1992-Jun 2002. Source FRED database: Monthly new houses started (HSTART), new single-family houses sold (HSOLD), single-family housing units completed (HCOMPL), trade balance (TBAL), disposable income (DINC), and durable goods (DGOOD). Variables are NSA. Data scaled between 0 and 1 and sensitivities sum to 1. Network Architecture: 6 input nodes, 3 hidden layer nodes, 1 output node.

Hypothesis 8. *Housing prices (SA), micro supply and demand factors. For the Case-Shiller SA test, housing prices are explained by new single-family houses sold, single-family housing completed, single-family housing started, disposable income, durable goods, and trade balance (demand by foreigners). Data extends from 1992 through June 2022.*

Hypothesis 9. *Housing prices and expectations. This hypothesis is motivated by Piazzesi and Schneider (2016) and Kuchler et al. (2022) on housing expectation as measured by the Michigan Consumer Sentiment Index and its numerous components. H9 adds four new variables: housing units authorized but not started, single-family units started/population level, Michigan consumer sentiment, and the Kansas City financial stress index. We find significant improvement based on a decline of mean absolute error versus H7. See Figure 9.*

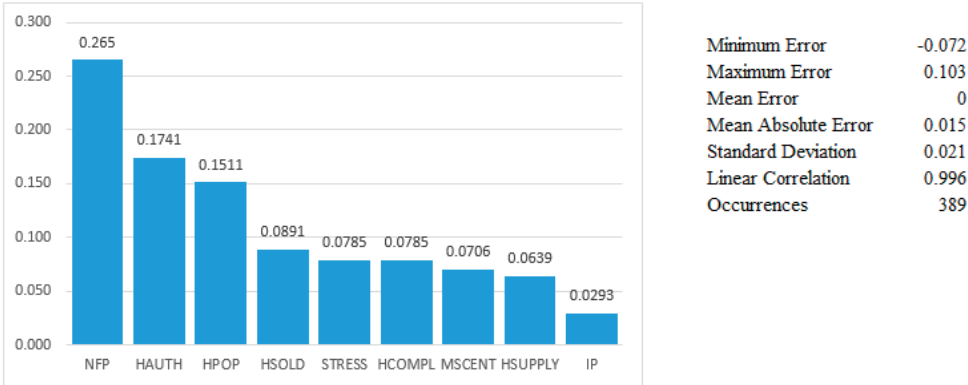


Figure 9. H9. Housing Prices micro supply and demand factors with expectations. Period: Feb 1990-June 2022. Source FRED database: Monthly non-farm payroll (NFP), housing units authorized but not started (HAUTH), single-family units started divided by the population level (HPOP), single-family houses sold (HSOLD), Kansas City financial stress index (STRESS), single-family housing units completed (HCOMPL), Michigan consumer sentiment (MSCENT), housing supply (HSUPPLY) and industrial production (IP). Variables are NSA. Data scaled between 0 and 1 and sensitivities sum to 1. Network Architecture: 9 input nodes, 5 hidden layer nodes, 1 output node.

We next test the importance of past prices as a driver for housing price behavior relative to key exogenous macro and micro variables. If key macro variables represent most of the variation in housing prices, then the bubble question becomes more complex. Housing prices are responding to changes in the business cycle, monetary policy, and sector-specific variables. Housing is cyclical, but these cycles may not demonstrate bubble behavior. Alternatively, if extrapolative behavior explains most of the variation in housing prices, then we can say that a positive feedback loop from past trends drives price extremes and dominates any exogenous variables. This hypothesis is based on momentum and bubble work developed by Piazzesi and Schneider (2016), Piazzesi et al. (2020), Mayer (2011), Glaeser and Gyourko (2018), and other surveys. Basco and Schäfer-i-Paradís (2022) follow a different methodology by proposing a model-free test of rational bubbles applied to US housing and conclude that a bubble occurred only during 2002–06.

Hypothesis 10. *Price extrapolation as the key driver of housing prices. This hypothesis uses all NSA data: Case–Shiller lagged 1 period, Case–Shiller lagged 6 periods, with Fed funds, housing units authorized but not started, single-family housing completed, Chicago Fed financial conditions, Michigan consumer sentiment, and Michigan inflation expectations.*

Beyond short-term (1-lag) and longer-term (6-lag) momentum, we include variables for housing expectations, inflation expectations, indicators such as housing permits for housing not started, completed houses that may reduce or fuel current prices, financial conditions evaluated in surveys, and other factors. The results in Figure 10 show the importance of extrapolative expectations as a key driver of housing market behavior.

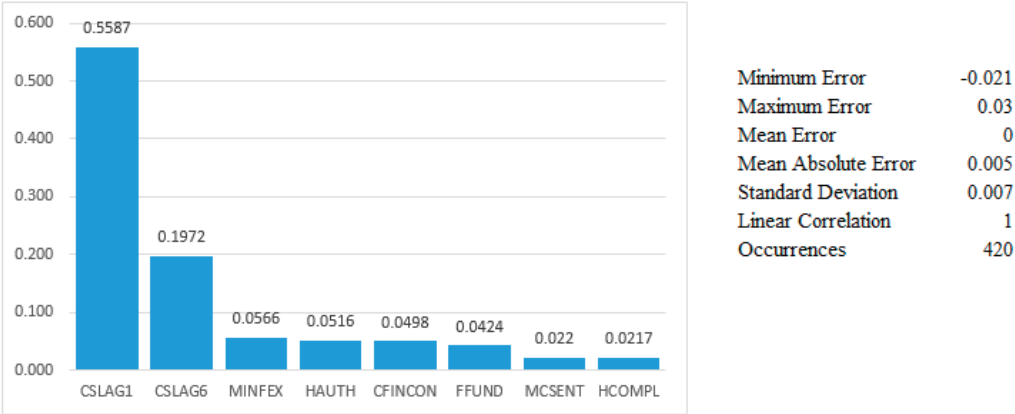


Figure 10. H10. Housing prices are driven by lagged prices. Period Jan 1992–June 2022. Source FRED database: Case–Shiller prices lagged 1 period (CSLAG1), Case–Shiller prices lagged 6 periods (CSLAG6), Michigan inflation expectations (MINFEX), housing units authorized but not started (HAUTH), single-family housing units completed (HCOMPL), Chicago Fed Financial conditions (CFINCON), Fed funds (FFUND), Michigan consumer sentiment (MCSENT), and single-family housing units completed (HCOMPL). Data scaled between 0 and 1 and sensitivities sum to 1. Network Architecture: 8 input nodes, 6 hidden layer nodes, 1 output node.

The past price components with lag 1 and lag 6 dominate and explain about 80% of the housing price variation. Other than price lag effects, the other two important variables are inflation expectations and housing units authorized but not started, followed by Fed funds and financial conditions.

In summary, the most impactful variables using non-seasonally adjusted data are consumer price index rent, the 30-year mortgage rate, Fed assets, non-farm employment, and the lag of the Case–Shiller index. CPI Rent was the top variable in H1. CSLag1 was

the top variable in H10, and Fed Assets was the top variable in H4, H5, and H6. The remaining hypotheses with NSA data (H3, H7, and H9) had either non-farm employment or the 30-year mortgage rate as the best independent variable. Running a single neural network on the most impactful non-seasonally adjusted variables, excluding the Case–Shiller lag, Fed assets is the most influential variable, as noted in H3, H5, and H6. For the hypotheses that use only seasonally adjusted data, H2 and H8, durable goods spending was the dominant variable.

Comparison of Results and Bubble Tests

Themes are presented in Table 2 with the first six hypotheses using NSA vs. SA data and the remaining four only using NSA data. The themes have different explanatory power based on the mean absolute error, the maximum and minimum error, and the standard error. The price extrapolation theme, which suggests the potential for a bubble through price expectations, dominates all models even with a restricted set of features. On a relative basis, the housing-specific information theme outperforms those that focus on macro variables, the business cycle, monetary policy, and the savings glut; yet, when testing over a shorter sample, we find the possibility of a bubble break for four hypotheses.

Table 2. Comparison of themes using neural network models.

Themes:	Housing and the Business Cycle		Housing Driven by Monetary Policy		Housing and Global Savings Glut		Housing Supply and Demand		Housing and Price	
	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10
Max Error	0.106	0.06	0.256	0.102	0.152	0.172	0.166	0.085	0.103	0.03
Min Error	−0.122	−0.072	−0.199	−0.083	−0.085	−0.125	−0.172	−0.044	−0.072	−0.021
Mean Abs Error	0.028	0.013	0.044	0.017	0.022	0.029	0.026	0.012	0.015	0.005
Std Dev	0.04	0.018	0.062	0.023	0.028	0.038	0.038	0.016	0.021	0.007
CUSUM test for each H; entire sample	reject *	accept	reject **	accept	accept	accept	accept	accept	accept	accept
CUSUM test for same sample Feb 2003–Jun 2022	reject *	reject *	reject **	accept	accept	accept	reject *	reject *	accept	accept
Bubble period above 0.05 critical level	5 Oct–22 Mar	7 Mar–8 Feb	None	None	None	None	4 May–6 Sep	7 Nov–9 Sep	None	None
Number of months above 0.05 critical level	78 months	12 months	None	None	None	None	29 months	19 months	None	None

* reject at 0.05 level, ** reject at 0.10 level, of no structural break (bubble) in the model.

The investigation of the first theme clearly shows that housing is closely tied to business cycles. Among various NSA macro variables, CPI rent and industrial production are the dominant drivers of housing variation, while durable goods and disposable income are the most important SA macro variables.

An analysis of the role of credit and monetary policy shows that among the NSA variables, the 30-year mortgage rate and the 10-year 2-year Treasury rate spread are the leading variables. However, the introduction of Fed assets after the GFC overtakes the 30-year mortgage rate as the key explanation for housing prices. The extensive availability of money and credit drove the surge in housing prices over the last decade. When monetary policy is compared with the global saving glut, SA factors such as the trade balance and disposable income still emerge as important housing price drivers.

Housing demand fundamentals such as durable goods and disposable income drive housing demand among SA variables. Non-farm employment, houses completed, and houses sold emerge among NSA variables as important features. New privately owned housing units scaled to the population level become an important measure of speculation; as housing prices increase, the number of adjusted single-family units also increases. Among the several financial variables considered, housing is closely associated with Fed funds and Fed assets.

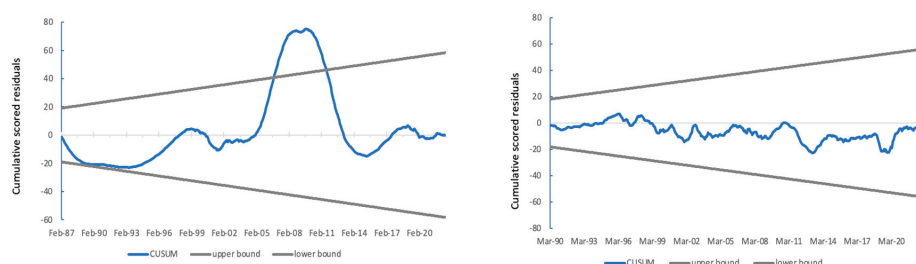
If, as some authors claim, there is a persistent deficit of several million houses in the US, the supply of housing is rather inelastic; however, we find good macro conditions, a global saving glut, and easy monetary policies drive housing price predictions over the past 25 years. Nonetheless, the University of Michigan sentiment index and the Kansas City financial condition stress index do not play any special role in housing price behavior. Lastly, using 1- and 6-period lags in the house price index to investigate the role of extrapolative forecasts, we find strong evidence that both lags dominate other features. Past prices play an important role in explaining the variation in housing prices. Extrapolative expectations fueled by quantitative easing, which provides excess credit, along with a savings glut story, cause more extreme home price changes. While all themes generate some validity, the dominance of extrapolative expectations with easy credit provides a theme consistent with bubble discussions.

To identify possible bubbles, we test for error dislocations or structural breaks, from the upper and lower band confidence regions for our ten model hypotheses. Although difficult to formally conclude given the problem of bubbles being a joint hypothesis, the residual extremes, which indicate a structural break, likely represent a bubble. The tests for whether residuals stay within the CUSUM bounds for all themes are reported for two different samples in Table 2.

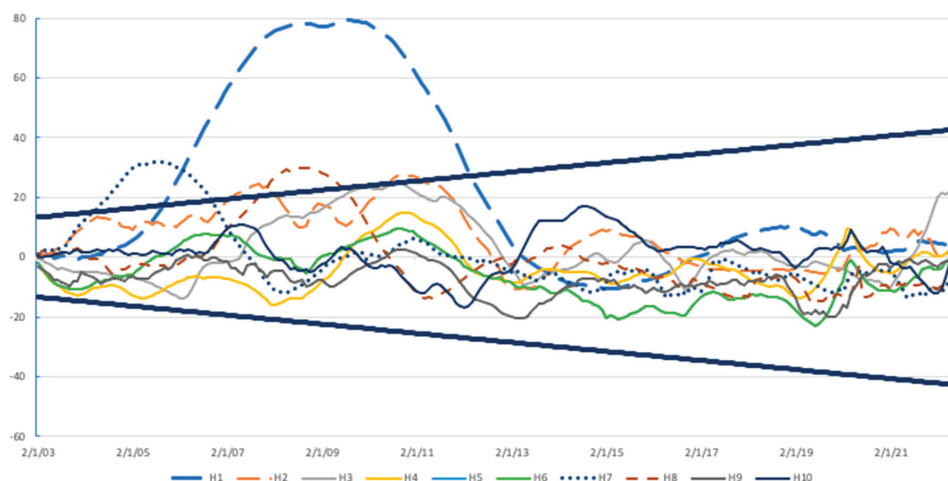
First, we test the null hypothesis of staying within the confidence band for the entire test period for each model. The sample size for each model test may differ, which means the confidence bounds will also change. In the case of a full sample for each model, we find two hypotheses reject the null of no structural break. However, in the case of H3, the rejection is caused by the CUSUM piercing the lower bound during the period prior to the tech bubble. That is, housing prices were too low relative to the monetary policy indicator model. In the case of H1, the CUSUM was consistent with a bubble prior to the GFC.

Second, we provide a set of tests where the start and end date are the same for all ten hypotheses and find the presence of potential bubbles in four cases. We include the start and end date of each break as well as the number of months the CUSUM of scored residuals (price > model) exceeded the upper confidence band. Although the timeframes for the possible presence of a break (bubble) differ in our four cases, all are consistent with the view that a housing bubble as defined by the CUSUM extremes, nonstationary mean or feature behavior, occurred prior to the Great Financial Crisis. Nonetheless, our tests also demonstrate that several models do not show the presence of any structural breaks and can thus explain the housing price extremes.

Bubble identification through structural breaks is highly model- or feature-dependent. For illustrative purposes, we show the CUSUM for two themes, H1 and H9. For H1, there is a clear positive CUSUM, actual greater than predicted, for the period prior to the GFC and then a switch to negative after the GFC, actual less than predicted, followed by a return to normal. See Scheme 1. In Scheme 2, the CUSUMs for each of the ten hypotheses are presented with the upper and lower bound for the average number of features with a start date when all data are available, February 2003 to June 2022. Note that the critical levels are based on the sample size and will be sensitive to the start date, and CUSUM tests will suffer from low power if the change point or bubble dislocation is close to the beginning or end of the sample.



Scheme 1. CUSUM chart examples for H1 and H9 with upper and lower bounds for critical 0.05 level.



Scheme 2. CUSUM for with upper and lower bounds for the 0.05 critical level with matching start date of February 2003.

This comparison shows a clear pattern for some models to reject the null hypothesis through piercing the upper confidence band only to return to normal. Notably, the SA and NSA business cycle models suggest a break, as well as the housing supply and demand models. Price extremes during the period of the perceived housing bubbles seem to be linked with monetary policy, the saving glut, and price extrapolation. In no case do we see extreme CUSUMs after the COVID-19 pandemic.

Different themes will have different CUSUM patterns as seen in Scheme 2, which suggest the possibility of structural breaks and misspecification of model narratives. For example, our business cycle theme shows that standardized residuals exceeded the predicted value only to then fall after the crisis. In this case, the macro narrative of just looking at the business cycle would lead to the belief in a housing bubble. The CUSUM pattern of a break and possible bubble is not found when we use an extrapolative price model or in our other macro hypotheses, especially when we account for the behavior of monetary policy.

5. Conclusions

We selected important features that can explain the long-run behavior of housing prices from January 1987 to June 2022, the emergence of the 2004–07 housing bubble, its bursting that led to the GFC, and the housing boom surrounding the COVID-19 pandemic. From these features, we developed ten hypotheses to study the US housing market for both SA and NSA house prices data over similar time periods and methodology. While the strong run-up in housing prices during COVID-19 may suggest a bubble, macro factors

like easy fiscal and monetary policies stimulated housing prices and prevented a possible strong reversal. There is no evidence based on our CUSUM analysis to suggest a COVID-19 housing bubble. Additionally, without a large correction following the rapid rise of house prices during COVID-19, we cannot argue that housing has experienced a second bubble during 1987–2022.

Our NN methodology recognizes the relative value of certain macro drivers such as the Consumer Price Index Rent, the 30-Year mortgage rate, Fed assets, and non-farm payroll for explaining the NSA S&P Case–Shiller US National Home Price Index. When the NN methodology is applied to the SA National Home Price Index, durable goods, Fed assets, new single-family houses sold, and disposable income emerge as important features. Among financial variables, mortgage rates, Fed funds and Fed assets are key drivers of house prices.

Tests for housing bubbles (structural breaks) are applied to our NN predictions and find that some hypotheses suggest there were price deviations consistent with housing bubbles. The hypothesis that focuses on extrapolative prices does not show the presence of a bubble (structural break) but confirms that the formation of expectations based on past prices is a key housing market driver.⁴ We do not dispute the results of others who find the presence of price extremes or bubbles; rather, we highlight the difficulty of identifying bubbles when there are competing hypotheses. A micro fundamental or price-based model may identify a bubble, yet a macro model may conclude that there are fundamental drivers that lead to high housing prices consistent with model parameters.

Any bubble is conditional on a model and the features used to describe the market environment. Comparing different hypotheses or housing narratives with an unstructured machine learning methodology using over 30 independent variables and explicit residual tests serves as an effective framework for broadly analyzing housing market extremes and bubbles. Using a single NN methodology applied to thematic model specifications coupled with tests for structural breaks (bubbles) advances our understanding of the housing market and supports the identification of different market themes. One man’s bubble could be another man’s rational behavior.

Author Contributions: All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by A.G.M., M.M. and M.S.R. All authors commented on all versions of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: All data available from the St. Louis FRED database. Variables used are listed in the Appendix A.

Conflicts of Interest: Mark S. Rzepczynski was employed by the company AMPHI Research and Trading, LLC. The authors have no sources of funding or conflicting interests.

Appendix A. Data from the St. Louis FRED Database (Accessed 25 October 2022)

Type	FRED Identifier	Description	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10
NSA	COMPUTNSA	New Privately Owned Housing Units Completed: Total Units, Thousands of Units, Monthly, Not Seasonally Adjusted							x		x	x
NSA	CSUSHPINSA	S&P/Case–Shiller U.S. National Home Price Index, Index Jan 2000 = 100, Monthly, Not Seasonally Adjusted	x		x	x	x		x		x	x
NSA	UMCSENT	University of Michigan: Consumer Sentiment, Index 1966: Q1 = 100, Monthly, Not Seasonally Adjusted									x	x

Type	FRED Identifier	Description	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10
NSA	AUTHNOTT	New Privately Owned Housing Units Authorized but Not Started: Total Units, Thousands of Units, Monthly, Not Seasonally Adjusted									x	x
NSA	MICH	University of Michigan: Inflation Expectation, Percent, Monthly, Not Seasonally Adjusted										x
NSA	FEDFUNDS	Federal Funds Effective Rate, Percent, Monthly, Not Seasonally Adjusted			x	x	x	x				x
NSA	IPB50001N	Industrial Production: Total Index, Index 2017 = 100, Monthly, Not Seasonally Adjusted	x						x		x	
NSA	KCFSI	Kansas City Financial Stress Index, Index, Monthly, Not Seasonally Adjusted									x	
NSA	MSPNHSUS	Median Sales Price for New Houses Sold in the United States, Dollars, Monthly, Not Seasonally Adjusted							x		x	
NSA	PAYNSA	All Employees, Total Nonfarm, Thousands of Persons, Monthly, Not Seasonally Adjusted	x						x		x	
NSA	HOUST1FNSA	New Privately Owned Housing Units Started: Single-Family Units, Thousands of Units, Monthly, Not Seasonally Adjusted									x	
NSA	CNP16OV	Population Level, Thousands of Persons, Monthly, Not Seasonally Adjusted									x	
NSA	CUUR0000SEHA	Consumer Price Index for All Urban Consumers: Rent of Primary Residence in U.S. City Average, Index 1982–1984 = 100, Monthly, Not Seasonally Adjusted	x									
NSA	EMVMACRO-CONSUME	Equity Market Volatility Tracker: Macroeconomic News and Outlook: Consumer Spending and Sentiment, Index, Monthly, Not Seasonally Adjusted	x									
NSA	EPUMONETARY	Economic Policy Uncertainty Index: Categorical Index: Monetary policy, Index, Monthly, Not Seasonally Adjusted	x		x	x						
NSA	MORTGAGE30US	30-Year Fixed Rate Mortgage Average in the United States, Percent, Monthly, Not Seasonally Adjusted	x		x	x	x	x				
NSA	MSACSRNSA	Monthly Supply of New Houses in the United States, Months' Supply, Monthly, Not Seasonally Adjusted							x		x	
NSA	NFCI	Chicago Fed National Financial Conditions Index, Index, Monthly, Not Seasonally Adjusted										x
NSA	T10Y2YM	10-Year Treasury Constant Maturity Minus 2-Year Treasury Constant Maturity, Percent, Monthly, Not Seasonally Adjusted			x	x						
NSA	USACPIALL-MINMEI	Consumer Price Index: All Items for the United States, Index 2015 = 100, Monthly, Not Seasonally Adjusted	x									

Type	FRED Identifier	Description	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10
NSA	WALCL	Assets: Total Assets: Total Assets (Less Eliminations from Consolidation): Wednesday Level, Millions of U.S. Dollars, Monthly, Not Seasonally Adjusted			x		x	x				
SA	BOPGSTB	Trade Balance: Goods and Services, Balance of Payments Basis, Millions of Dollars, Monthly, Seasonally Adjusted		x			x	x		x		
SA	COMPU1USA	New Privately Owned Housing Units Completed: Single-Family Units, Thousands of Units, Monthly, Seasonally Adjusted Annual Rate									x	
SA	CSUSHPISA	S&P/Case-Shiller U.S. National Home Price Index, Index Jan 2000 = 100, Monthly, Seasonally Adjusted		x				x		x		
SA	DSPI	Disposable Personal Income, Billions of Dollars, Monthly, Seasonally Adjusted Annual Rate		x							x	
SA	HOUST1F	New Privately Owned Housing Units Started: Single-Family Units, Thousands of Units, Monthly, Seasonally Adjusted Annual Rate									x	
SA	HSN1F	New One Family Houses Sold: United States, Thousands, Monthly, Seasonally Adjusted Annual Rate									x	
SA	UNEMPLOY	Unemployment Level, Thousands of Persons, Monthly, Seasonally Adjusted		x								
SA	PCEDG	Personal Consumption Expenditures: Durable Goods, Billions of Dollars, Monthly, Seasonally Adjusted Annual Rate		x							x	
SA	EMRATIO	Employment-Population Ratio, Percent, Monthly, Seasonally Adjusted		x								
SA	INDPRO	Industrial Production: Total Index, Index 2017 = 100, Monthly, Seasonally Adjusted		x								

- Notes**
- ¹ See the Dallas Fed’s house price database, which updates quarterly exuberance indicators based on price-to-fundamentals and price-to-rent ratios, <https://www.dallasfed.org/research/international/houseprice> (accessed on 9 August 2023).
- ² Machine learning has been applied to many housing market studies; see Xu and Zhang (2021) for a review of this work. Since bubble identification is a joint hypothesis, our approach is to use a model-free approach focused on using specific features associated with a macro narrative as a base for comparison.
- ³ We employ IBM Modeler; for details, see the Modeler 18 Algorithms Guide, p. 311.
- ⁴ A comparison of our macro results with the micro exuberance indicators using supremum augmented Dickey–Fuller (SADF) and generalized SADF (GSADF) reported by the Dallas Fed in their International House Price Database suggests that a macro model may provide useful information on the housing price extreme drivers. The SADF test shows exuberance or a bubble before the GFC based on price or price-to-fundamental ratios; however, it does not provide any insight on macro drivers or macro narratives that can explain a bubble.

References

Basco, Sergi, and Maximilian Schäfer-i-Paradís. 2022. A Model-Free Test of Rational Bubbles: An Application to the US Housing Market. Available online: <https://ssrn.com/abstract=4161754> (accessed on 9 August 2023).

- Bernanke, Ben. 2005. The Global Saving Glut and the U.S. Current Account Deficit. Paper presented at Remarks at the Sandridge Lecture, Virginia Association of Economics, Richmond, VA, USA, March 10. Available online: <http://www.federalreserve.gov/boarddocs/speeches/2005/200503102/> (accessed on 9 August 2023).
- Bernanke, Ben S. 2007a. Global Imbalances: Recent Developments and Prospects. Paper presented at the Bundesbank Lecture, Berlin, Germany, September 11.
- Bernanke, Ben S. 2007b. Globalization and Monetary Policy. Paper presented at the Fourth Economic Summit, Stanford Institute for Economic Policy Research, Stanford, CA, USA, March 2.
- Bernanke, Ben S. 2009. Asia and the Global Financial Crisis. Paper presented at the Federal Reserve Bank of San Francisco's Conference on Asia and the Global Financial Crisis, Santa Barbara, CA, USA, October 19.
- Bernanke, Ben S. 2010. Monetary Policy and the Housing Bubble. Paper presented at the Annual Meetings of the American Economic Association, Atlanta, Georgia, January 3. Available online: <https://www.federalreserve.gov/newsevents/speech/bernanke20100103a.htm> (accessed on 9 August 2023).
- Bhar, Ramaprasad, and Anastasios George Malliaris. 2021. Modeling U.S. Monetary Policy During the Great Financial Crisis and Lessons for COVID-19. *Journal of Policy Modeling* 43: 15–33. [CrossRef] [PubMed]
- Bhar, Ramaprasad, Mary Malliaris, Mark Rzepczynski, and A. G. Malliaris. 2024. Five Themes of U.S. Home Prices Cycles: A Dynamic Modelling Approach. *Annals of Operations Research*. [CrossRef]
- Corsi, Fulvio, and Didier Sornette. 2014. Follow the Money: The Monetary Roots of Bubbles and Crashes. *The International Review of Financial Analysis* 32: 47–59.
- Duca, John, John Muellbauer, and Anthony Murphy. 2021. What Drives House Price Cycles? *International Experiences and Policy Issues Journal of Economic Literature* 59: 773–864.
- Erdmann, Kevin. 2023. Reassessing the Role of Supply and Demand on Housing Bubble Prices. Available online: <https://ssrn.com/abstract=4327414> (accessed on 9 August 2023).
- Evanoff, Douglas D., George G. Kaufman, and Anastasios G. Malliaris, eds. 2012. *New Perspectives on Asset Price Bubbles*. New York: Oxford University Press.
- Evgenidis, Anastasios, and Anastasios Malliaris. 2023. House Bubbles, Global Imbalances and Monetary Policy in the U.S. *Journal of International Money and Finance* 138: 102919. [CrossRef]
- Glaeser, Edward, and Joseph Gyourko. 2018. The Economic Implications of Housing Supply. *Journal of Economic Perspectives* 32: 3–30. [CrossRef]
- Haykin, Simon. 1994. *Neural Networks: A Comprehensive Foundation*. Englewood Cliffs: Macmillan Publishing Company.
- Holmes, Mark J., Jesús Otero, and Theodore Panagiotidis. 2011. Investigating regional house price convergence in the United States: Evidence from a pair-wise approach. *Economic Modelling* 28: 2369–76. [CrossRef]
- Homm, Ulrich, and Jorg Breitung. 2012. Testing for Speculative Bubbles in Stock Markets: A Comparison of Alternative Methods. *Journal of Financial Econometrics* 10: 198–231. [CrossRef]
- Hu, Yang. 2023. A Review of Phillips-type Right-tailed Unit Root Bubble Detection Tests. *Journal of Economic Surveys* 37: 141–58. [CrossRef]
- Kindleberger, Charles. 1996. *Manias, Panics and Crashes: A History of Financial Crises*, 3rd ed. New York: John Wiley and Sons.
- Kiselev, Alexander, and Lenya Ryzhik. 2010. A simple model for asset price bubble formation and collapse. *arXiv arXiv:1009.0299*.
- Kuchler, Theresa, Monika Piazzesi, and Johannes Stroebel. 2022. Housing Market Expectations. In *Handbook of Economic Expectations*. Cambridge: Academic Press.
- Leamer, Edward E. 2015. Housing Really is the Business Cycles: What Survives the Lessons of 2008-09? *Journal of Money, Credit and Banking* 47: 43–50. [CrossRef]
- LeRoy, Stephen F. 2012. Infinite Portfolio Strategies. *Contemporary Economics* 6: 54–60. [CrossRef]
- Levitin, Adam J., Desen Lin, and Susan M. Wachter. 2020. Mortgage Risk Premiums during the Housing Bubble. *The Journal of Real Estate Finance and Economics* 60: 421–68. [CrossRef]
- Malliaris, Anastasios, and Mary Malliaris. 2021. What Microeconomic Fundamentals Drove Global Oil Prices during 1986–2020? *Journal of Risk and Financial Management* 14: 391. [CrossRef]
- Martinez-Garcia, Enrique, and Valerie Grossman. 2018. Explosive Dynamics in House Prices? An Exploration of Financial Market Spillovers in Housing Markets around the World. *Journal of International Money and Finance* 101: 102103. [CrossRef]
- Mayer, Christopher J. 2011. Housing Bubbles: A Survey. *Annual Review of Economics* 3: 559–77. Available online: <https://ssrn.com/abstract=1920051> (accessed on 9 August 2023). [CrossRef]
- McDonald, John F., and Houston H. Stokes. 2013. Monetary Policy and the Housing Bubble. *Journal of Real Estate Finance and Economics* 46: 437–51. [CrossRef]
- Otero, Jesús, Theodore Panagiotidis, and Georgios Papapanagiotou. 2022. Testing for exuberance in house prices using data sampled at different frequencies. *Studies in Nonlinear Dynamics & Econometrics* 26: 675–91.
- Phillips, Peter C. B., and Shuping Shi. 2018. Real Time Monitoring of Asset Markets: Bubbles and Crises. In *Handbook of Statistics*. Cowles Foundation Discussion Paper No 2152. Amsterdam: Elsevier.
- Piazzesi, Monika, and Martin Schneider. 2016. *Housing and Macroeconomics*. NBER Working Paper 22354. Chicago: University of Chicago Press.

- Piazzesi, Monika, Martin Schneider, and Johannes Strobel. 2020. Segmented Housing Search. *American Economic Review* 110: 720–59. [CrossRef]
- Ploberger, Werner, and Walter Krämer. 1992. The CUSUM test with OLS residuals. *Econometrica* 60: 271–85. [CrossRef]
- Sa, Filipa, and Tomasz Wieladex. 2015. Capital Inflows and U.S. Housing Boom. *Journal of Money, Credit and Banking* 47: 221–56. [CrossRef]
- Saltelli, Andrea, Stefano Tarantola, Francesca Campolongo, and Marco Ratto. 2004. *Sensitivity Analysis in Practice—A Guide to Assessing Scientific Models*. Hoboken: John Wiley & Sons Publishing.
- Scherbina, Anna, and Bernd Schlusche. 2014. Asset price bubbles: A survey. *Quantitative Finance* 14: 589–604. [CrossRef]
- Shi, Shuping, and Peter C. B. Phillips. 2020. Diagnosing Housing Fever with An Econometric Thermometer. *Journal of Economic Surveys* 37: 159–86. [CrossRef]
- Sornette, Didier, and Peter Cauwels. 2014. Financial Bubbles: Mechanisms and Diagnostics. *arXiv* arXiv:1404.2140. [CrossRef]
- Sornette, Didier. 2003. *Why Stock Markets Crash: Critical Events in Complex Financial Systems*. Princeton: Princeton University Press.
- Sornette, Didier, Peter Cauwels, and Georgi Smilyanov. 2017. Can We Use Volatility to Diagnose Financial Bubbles? Lessons from 40 Historical Bubbles. *Quantitative Finance and Economics* 2: 486–594. Available online: <https://ssrn.com/abstract=3006642> (accessed on 9 August 2023). [CrossRef]
- Steinberg, Joseph. 2019. On the Source of the U.S. Trade Deficits: Global Saving Glut or Domestic Saving Drought? *Review of Economic Dynamics* 31: 200–23. [CrossRef]
- Tarlie, Martin B., Georgios Sakoulis, and Roy Henriksson. 2022. Stock Market Bubbles and Anti-Bubbles. *International Review of Financial Analysis* 81: 101235. [CrossRef]
- Taylor, John B. 2007. Housing and Monetary Policy. In *Federal Reserve Bank of Kansas City, Housing, Housing Finance and Monetary Policy*. Kansas City: Federal Reserve Bank of Kansas City, pp. 463–76.
- Taylor, John B. 2009. *The Financial Crisis and the Policy Responses: An Empirical Analysis of What Went Wrong*. Working Paper, No.14631. Cambridge: National Bureau of Economic Research.
- Taylor, John B. 2010. The Fed and the Crisis: A Reply to Ben Bernanke. *Wall Street Journal* 11: A19.
- West, Kenneth D. 1988. Bubbles, Fads and Stock Price Volatility Tests: A Partial Evaluation. *The Journal of Finance* 43: 639–56. [CrossRef]
- Xu, Xiaojie, and Yun Zhang. 2021. Rent Index Forecasting Through Neural Networks. *Journal of Economic Studies* 49: 1321–39. [CrossRef]
- Ziemba, William, Sebastien Lleo, and Mikhail Zhitlukhin. 2017. Other Bubble-testing Methodologies and Historical Bubbles. World Scientific Book Chapters. In *Stock Market Crashes Predictable and Unpredictable and What to Do about Them*. Singapore: World Scientific Publishing Company, chp. 11, pp. 247–57.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

May 2024 Buy-Sell Guide for Dow Jones 30 Stocks and Modified Omega Criterion

H. D. Vinod

Institute for Ethics and Economic Policy (IEEP), Fordham University, Bronx, New York, NY 10458, USA; vinod@fordham.edu

Abstract: We study recent monthly data to help long-term investors buy or sell from the 30 Dow Jones Industrial Average (DJIA) Index components. The recommendations are based on six stock-picking algorithms and their average ranks. We explain the reasons for ignoring the claim that the Sharpe ratio algorithm lacks monotonicity. Since the version of “omega” in the literature uses weights that distort the actual gain–pain ratio faced by investors, we propose new weights. We use data from 30 stocks using the past 474 months (39+ years) of monthly closing prices, ending in May 2024. Our buy-sell recommendations also use newer “pandemic-proof” out-of-sample portfolio performance comparisons from the R package ‘generalCorr’. We report twelve sets of ranks for both out-of- and in-sample versions of the six algorithms. Averaging the twelve sets yields the top and bottom k stocks. For example, $k = 2$ suggests buying Visa Inc. and Johnson & Johnson while selling Coca-Cola and Procter & Gamble.

Keywords: Sharpe ratio monotonicity; stock picking tools; out-of-sample performance; pandemic-proof

1. Introduction

Investing in the stock market is vital in directing national resources to the most productive applications. The stock-picking activity of the financial services sector annually commands billions of dollars in fees. Malkiel (2013) argues that ever-growing fees for financial services are a deadweight loss for the US economy. The misallocation of resources creates waste and other losses to the economy. This paper offers tools to reduce such losses by proposing public domain stock-picking tools using free, open-source software by R Core Team (2023).

Wall Street investment outfits use stock prices attached to stock symbols. We use pretty long price data concerning 39 years and six months, ending in May 2024. If \$1 is invested in buying a stock priced at P_t at time t , if the price (adjusted for dividends) at time $t + 1$ is higher, the net return $r_t = [(P_{t+1} - P_t)/P_t]$ will exceed the initial investment of \$1. Since the net return is negative when losses are incurred, one defines the gross return as $(1 + r_t) = 1 + (P_{t+1} - P_t)/P_t = P_{t+1}/P_t$. The gross return is always positive since prices are positive.

Continuously compounded return is the exponential return, $\exp(r_t)$, which is always positive. The series expansion of $\exp(r_t)$ is $(1 + r_t + r_t^2/2! + \dots)$. Assuming higher order terms in the expansion can be ignored, $(|r_t| < 1)$, one can write $\exp(r_t) = 1 + r_t$. It is customary to equate the exponential return to the gross return and write $r_t = \log(P_{t+1}/P_t) = \log P_{t+1} - \log P_t$. Many published papers use the *first difference* of logs of prices evaluated at time $t + 1$ as a return from investment. Since the return data do not always satisfy $(|r_t| < 1)$ for certain time periods, let us define $r_t = [(P_t - P_{t-1})/P_{t-1}]$ for our monthly returns from the Dow Jones Industrial Average’s 30 (DJ30) stocks.

Let $f(r)$ denote the probability distribution function of returns (r) from investing in one of the 30 stocks, and let $F(r)$ denote the (cumulative) distribution function of returns. Denote the expected value (mean) of $f(r)$ by μ and the standard deviation by σ . The Sharpe

Citation: Vinod, H. D. 2024. May 2024 Buy-Sell Guide for Dow Jones 30 Stocks and Modified Omega Criterion. *Journal of Risk and Financial Management* 17: 343. <https://doi.org/10.3390/jrfm17080343>

Academic Editor: Thanasis Stengos

Received: 25 June 2024

Revised: 5 August 2024

Accepted: 6 August 2024

Published: 8 August 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

ratio, Sharpe (1966), is named after a Nobel-winning economist. It represents a risk-adjusted average return from investment X , defined as

$$SR(r_X) = \mu(r_X) / \sigma(r_X), \quad (1)$$

where the denominator makes risk adjustment. Typical SR ranks investment opportunities based on data on market returns represented by the probability distribution $f(r)$ mentioned before.

Although (1) can also be used for ranking gambles, this paper uses it for stock-picking. Gambling is a zero-sum or negative-sum game. By contrast, buying (selling) a share of companies contributing to socially desirable (undesirable) goods and services often yields positive μ and a larger SR.

Aumann and Serrano (2008) propose an ‘index of riskiness’ of an investment as if it is a gamble. They conjure imaginary gambling situations to assess their riskiness using an axiomatic framework and an economic decision-making context. They reject the reciprocal of SR of (1) as a measure of riskiness because it fails monotonicity property based on first-order stochastic dominance. Aumann and Serrano (2008) focus too much on the attributes of the stock buyer measured by a buyer’s risk-averse utility function. These authors refer to a constant absolute risk-averse “CARA person” (page 816).

Cheridito and Kromer (2013), or “CK13,” is an impressive study of mathematical formulas defining 45 performance measures similar to the reward–risk ratio (1). Among the 45 are tweaks on Sharpe ratios and ratios involving the value at risk (VaR). They evaluate the following four properties:

- **(M) monotonicity** means that more is better than less. This is a common-sense minimal requirement. All performance measures should satisfy it.
- **(Q) Quasi-concavity** describes uncertainty aversion linked to economists’ utility and decision theories. It is not relevant if we include investors who allocate a (small) proportion of investor funds as if they are risk-loving. We need not accept this criterion for our purposes.
- **(S) Scale Invariance** or $SR(X) = SR(\lambda X)$ where λ is a constant. The invariance requirement is inappropriate for modern investing where some technologies need very large (or small) scale and transaction costs are low for large transactions.
- **(D) Distribution-based**. All six stock-picking algorithms in this paper are distribution-based.

CK13 evaluations are not necessarily for stock market investment, but include the ranking of gambles where the “probability measure” (as in the measure theory of Statistics) may be unknown. By contrast, our probability measure is well approximated by $f(r)$, and its riskiness is well represented by its dispersion. CK13 claim that 17 measures out of 45, including the Sharpe ratio (SR), fail to meet the minimal monotonicity requirement **(M)**. The next four paragraphs show the limitations of the CK13 claim. We shall show that CK13 proof assumes certain artificial gambles that are irrelevant for our stock-picking based on $f(r)$.

Proposition 4.1 in CK13 allegedly proves that $SR = \mu(X) / \sigma(X)$ fails **(M)**. The proof involves the existence of pathological cases involving three constants satisfying properties as follows: $[(a, b, c) > 0, a > c, b \geq 1, bc > a]$. The proof needs to assume a non-negative random variable $Z \geq 0$. Since stock returns can be both positive and negative, the assumption $Z \geq 0$ is invalid for our $f(r)$.

We begin by showing that SR fails **(M)** when $Z \geq 0$. CK13 proof defines $X = a + b * Z$ and $Y = c + Z$. For example, choose $a = 2.5, b = 2.4, c = 1.5, Z = 1$, and $\sigma(Z) = 1$. Now, $X = 2.5 + 2.4 * Z = 4.9$ and $Y = 1.5 + Z = 2.5$, with $X > Y$. Now, $E(X) = 4.9$ is 96% larger than $E(Y) = 2.5$, whereas $\sigma(X) = 2.4$ is 240% larger than $\sigma(Y) = 1$. Hence, $SR(X) = 4.9/2.4 = 2.04 < SR(Y) = 2.5$. Since X is larger while $SR(X)$ is smaller, the Sharpe ratio indeed fails **(M)** for $Z \geq 0$.

Consider a realistic negative realization of the random variable Z , representing losses, with $Z = -1$. Let us keep the above choices of (a, b, c) and formulas for X and Y in CK13 unchanged. Now, $X = 2.5 + 2.4 * Z = 0.1$ and $Y = 1.5 + Z = 0.5$, implying that $X < Y$. This $E(X) = 0.1$ is 500% smaller than $E(Y) = 0.5$, whereas $\sigma(X) = 2.4$ is 240% larger than $\sigma(Y)=1$. Hence, $SR(X) = 0.1/2.4 = 0.04 < SR(Y) = 2.5$. Since X is smaller while $SR(X)$ is also smaller, the Sharpe ratio passes **(M)**.

The CK13 alleged failure to pass **(M)** holds only for artificially constructed gambles satisfying peculiarly demanding unrealistic restrictions ($X = a - bc + bY$) with a non-negative random variable Z denying any presence of losses, unlike typical $f(r)$. Thus, stock market investors can safely ignore the claim that SR fails **(M)**.

We further argue that property **(Q)** linked to utility theory can be ignored when allocating resources. After all, economists have long avoided interpersonal utility comparisons. An individual investor’s utility experience is personal, rarely identical across individuals, and exhibits marked change for the same individual over time and space. Psychologists have documented that utility from profits and losses is asymmetric and sensitive to profit and loss sizes. It is futile to assess whether someone is a “CARA person” before deciding which decision theory applies to him. In summary, the **M**, **Q**, **S**, and **D** criteria proposed by CK13 can be misleading when applied to stock-picking purposes.

Let us turn to the 30 stocks comprising DJIA. We begin with Tables 1 and 2 for the company names studied here in two parts with fifteen companies each. We report ticker symbols, relative weights in DJIA, and a (case-sensitive) single-character name to identify the stock for later use in graphics and tables. Figure 1 is inspired by Markowitz’s efficient frontier model, without the straight line representing a risk-free rate. This Figure has the mean return on the vertical axis and the standard deviation of returns measuring the volatility (risk) for that stock on the horizontal axis. We depict thirty letters in Figure 1 for each DJIA index component sticker.

The basic idea behind Figure 1 is that we imagine grouping stocks into a certain number (=7) of unequal width ranges of standard deviation class intervals. Our 30 stocks are implicitly assigned to these seven σ intervals. Now, the stock yielding the highest average return for each level of risk (measured by the midpoints of the sd class interval) dominates all those below it in Figure 1. The dominating stocks from DJIA are graphically identified as (j, v, h, f, C, a, z). The corresponding longer company names of dominating stocks, according to Tables 1 and 2, are Johnson & Johnson, Visa, Home Depot, Microsoft, Salesforce, Apple, and Amazon.

Table 1. Company names, ticker symbols, weight in DJIA, and abbreviations with only one character. Part 1.

Seq.	Company	Symbol	Weight	Char1
1	Apple Inc.	AAPL	2.93	a
2	Amgen Inc.	AMGN	4.65	A
3	Amazon.com Inc.	AMZN	2.99	z
4	American Express Co.	AXP	4.09	x
5	Boeing Co.	BA	2.88	b
6	Caterpillar Inc.	CAT	5.83	c
7	Salesforce Inc.	CRM	4.71	C
8	Cisco Systems Inc.	CSCO	0.83	S
9	Chevron Corp.	CVX	2.85	e
10	Walt Disney Co.	DIS	1.94	d
11	Dow Inc.	DOW	0.97	D
12	Goldman Sachs Group Inc.	GS	7.24	g
13	Home Depot Inc.	HD	5.72	h
14	Honeywell International Inc.	HON	3.33	H
15	Intl Business Machines Corp.	IBM	2.91	i

Table 2. Company names, ticker symbols, weight in DJIA, and abbreviations with only one character. Part 2.

Seq.	Company	Symbol	Weight	Char1
16	Intel Corp.	INTC	0.61	I
17	Johnson & Johnson	JNJ	2.53	j
18	JP Morgan Chase & Co	JPM	3.33	J
19	Coca Cola Co.	KO	1.06	k
20	McDonald's Corp.	MCD	4.75	m
21	3m Co.	MMM	1.58	M
22	Merck & Co. Inc.	MRK	2.25	K
23	Microsoft Corp.	MSFT	6.88	f
24	Nike Inc Cl B	NKE	1.62	n
25	Procter & Gamble Co.	PG	2.80	p
26	Travelers Cos Inc.	TRV	3.69	t
27	Unitedhealth Group Inc.	UNH	8.52	u
28	Visa Inc Class A Shares	V	4.75	v
29	Verizon Communications Inc.	VZ	0.68	V
30	Walmart Inc.	WMT	1.04	w

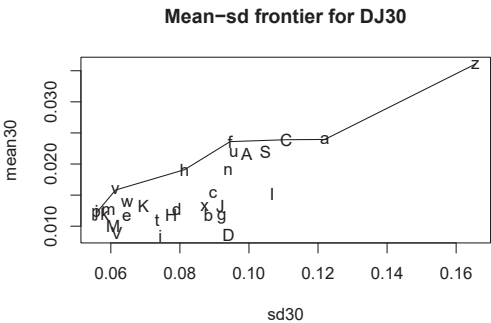


Figure 1. Mean–standard deviation efficiency frontier for Dow Jones 30 stocks.

1.1. Descriptive Statistics for the DJIA Stock Returns

This section reports some basic information about our data using standard descriptive stats. We report ‘min’ for the smallest return, Q1 for the first quartile, where 25% of data are below Q1 and 75% are above Q1. ‘Median’ and ‘Mean’ are self-explanatory. Q3 is for the third quartile (75% below and 25% above), and ‘max’ denotes the largest return. In finance, two additional descriptive stats are referenced. The $SR = \mu/\sigma$ defined before and the ‘expected gain–expected pain ratio’. The latter is called ‘omega’ (Ω) in Keating and Shadwick (2002), or “KS02” hereafter. While we include SR and Ω among the descriptive statistics associated with each stock, we include them among our six stock-picking algorithms.

1.1.1. Sharpe Ratios for Risk-Adjusted Stock Returns

The SR of (1) is a popular stock-picking tool. Many researchers have studied it and suggested modifications. SR treats a symmetric measure σ as an approximation to the risk. It treats volatility on the profit and loss side as equally undesirable. Volatility on the loss side is undesirable but desirable on the profit side. An adjusted version, $SR_{dsd} = \mu/DSD$, replaces the σ in the original denominator by downside standard deviation ($DSD = \sqrt{DSV}$) from the square root of DSV, the downside variance. Let n^b denote the number of observations below the mean. The downside variance in Vinod and Reagle (2005) (page 111, Equation (5.2.2)) is computed as

$$DSV = [\sum_t w_t (r_t - \bar{r})^2] / n^b, \tag{2}$$

where $w_t = 0$ when $r_t > \bar{r}$, and $w_t = 1$ otherwise. The summation is over n^b returns involving possible losses. The adjusted version

$$SR_{dsd} = \mu / \sqrt{DSV}, \quad (3)$$

is not difficult to compute. The online Appendix A provides software for SR_{dsd} .

Page 114 of Vinod and Reagle (2005) was perhaps the first to identify an often-ignored serious problem with Sharpe ratios. Now, we show how SR provides the wrong rankings when risk-adjusting any stocks with negative average returns. We explain the wrong ranking using an example of two stocks, p and q, with Sharpe ratios, $SR_p = \mu_p / \sigma_p$, and $SR_q = \mu_q / \sigma_q$. Now, assume that both have a negative average loss of 100, or $\mu_p = \mu_q = -100$. Next, assume that stock p is twice as volatile (risky) as stock q, or $\sigma_p = 10$ and $\sigma_q = 5$. We expect that the money-losing and more risky stock p is worse than q. Note that $SR_p = -100/10 = -10$ is larger than $SR_q = -100/5 = -20$. The ranking by SR says q is worse than p, which is against common sense.

Page 114 of Vinod and Reagle (2005) also explains how to use suitably large “add factors” to obtain the SR to yield the correct ranking. Fortunately, columns entitled ‘Mean’ in Tables 3 and 4 (of descriptive stats reported later) show that all thirty stocks have positive average returns. Thus, we do not need any ‘add factor’ adjustment in our context.

Note that our $SR = \mu / \sigma$ assumes that the researcher has true unknown values of μ and σ , rather than their sample estimates ignoring estimation errors. A “Double Sharpe ratio” divides the sample estimate $\hat{sr}$ by its standard error (SE), the standard deviation of the sampling distribution. The division penalizes stocks with a higher estimation risk. A double SR, which penalizes for estimation risk represented by a bootstrap standard error, is discussed on page 227 of Vinod and Reagle (2005). We denote it as SR with subscript ‘se’ as

$$SR_{se} = \hat{sr} / SE(sr), \quad (4)$$

The online Appendix A provides software for computing SR_{dsd} of (3) using downside standard deviation in the denominator. Next, the appendix software computes a large number J of SR_{dsd}^j based on bootstrap replicates. Then, the code computes $SE(sr)$ in the denominator of (4), which becomes the standard deviation of J bootstrap values SR_{dsd}^j .

Recall that the risk (horizontal axis) versus return (vertical axis) scatterplot of Figure 1 suggests that Johnson & Johnson, Visa, Home Depot, Microsoft, Salesforce, Apple, and Amazon graphically dominate others. The respective Sharpe ratios of these stocks are (0.22, 0.26, 0.23, 0.25, 0.22, 0.19, 0.15). Note that the Sharpe ratio is a direct measure of risk-adjusted return, bypassing the grouping of stock returns into standard deviation (sd) intervals.

The lowercase versions of ticker symbols for the top seven Sharpe ratios in increasing order of magnitude are amzn = 0.22, jnj = 0.22, crm = 0.22, unh = 0.23, hd = 0.23, msft = 0.25, and v = 0.26, respectively. Markowitz’s theory supports a stock-picking algorithm based on the Sharpe ratio. It is one of the six algorithms listed later in Section 3.

1.1.2. Original Computation of “Omega” for Stock Returns

KS02 name a “cumulative probability weighted” gain–loss ratio “omega” (Ω), and claim that it is a “universal” performance measure. We regard it as one of many and find that the weighting used in its original computation can be simplified without sacrificing its intent. This subsection describes the need for simplification, while the following subsection describes a recommended version. A referee has suggested that some readers are not interested in knowing exactly how our simplified version maintains the intent behind the gain–loss ratio. Uninterested readers can skip the present subsection.

The title of KS02 claims universality for their performance measure. Since a larger omega means a larger preponderance of positive returns, a stock having a larger omega is more desirable. The idea of measuring a gain–loss ratio is first mentioned by Bernardo and Ledoit (2000), who define r as a risk-adjusted excess return over a target return. Their risk adjustment requires assumptions about the utility function of investors. This paper avoids

any assumption regarding investor utility functions. We retain their distinction between $r^+ = \max(0, r)$, the positive part of returns, and $r^- = \max(r, 0)$, the negative part. Their gain–loss ratio is Ω_{bl} , (where the subscript ‘bl’ identifies authors) is as follows:

$$\Omega_{bl} = E(r^+) / E(r^-). \quad (5)$$

Section 3.1 of CK13 states that Bernardo–Ledoit-type ratios satisfy all four (**M**, **Q**, **S**, **D**) properties mentioned above.

KS02 do not assume anything about the utility function of investors. The gains and losses in KS02 are compared to a target return. If all stocks in a data set have a common target return, we can subtract it from all returns and work with returns in “excess of target”. Hence, there is no loss of generality in letting the target be zero. Therefore, our target return is mostly zero in the sequel. Metel et al. (2017) explain how omega generalizes the mean variance-based Sharpe ratio by encompassing the entire $f(r)$ (page 4) “therefore incorporating higher moment properties”. Their Theorem 1 proves that, within the class of elliptical density functions, using the Sharpe ratio or the Omega measure to optimize portfolio performance leads to the same optimal portfolio. The theorem has no practical implications for our purposes when working with real-world market returns, usually not from any known density.

KS02 define

$$\Omega_{ks} = E(\text{gain}) / E(\text{loss}), \quad (6)$$

where the expectation operator from the probability theory in $E(\text{gain})$ is represented by choosing weights $(1 - F(r))$ by assuming continuous distributions. Similarly, KS02 represent $E(\text{loss})$ by choosing $F(r)$ weights. The subscript ‘ks’ in Ω_{ks} identifies the authors of KS02.

Considerable literature on “omega” assumes a continuous $f(r)$, usually with a known (Gaussian, Elliptical) form. By contrast, this paper assumes a nonparametric representation of $f(r)$ based on actual market returns data. Since observable data are always discrete, we replace $F(r)$ with the empirical cumulative distribution function (ECDF), a step function of returns. It uses all the information in the data on returns and is considered a “sufficient statistic”.

Figure 2 illustrates the ECDF for an imaginary stock A with $n = 4$, returns $r_t = (-3, -1, 2, 5)$, for $t = 1, 2, \dots, n$. A vertical axis at $r = 0$, shown in Figure 2, separates the ECDF for stock A into the loss side on the left (for $r^- = \max(r, 0)$) and the gain side on the right (for $r^+ = \max(0, r)$).

How do we represent the E operator weights in the discrete case? Usually, mathematical expectation $E(x) = \sum x_t p_t$ (return x_t with probability p_t) is the average return. KS02 formulate the mathematical expectation of aggregate loss based on cumulative sums of negative returns r_t times corresponding cumulative probabilities from $F(r)$ as weights. Their gain side weights based on $(1 - F(r))$ are from the areas above the ECDF steps. The following paragraph uses Figure 2 to explain the numerical computation of KS02’s cumbersome weighting for toy stock A mentioned above.

The ECDF on the loss side for toy stock A has two negative ranges, (box intervals) $(-3, -1)$ and $(-1, 0)$, with areas under the pillars for the two ranges of $(1/n)$ and $(2/n)$, respectively. The respective gain side weights $(2/n)$ and $(1/n)$ are based on $(1 - F(r))$. These weights represent the areas above the pillars for the two ECDF ranges $(0, 2)$ and $(2, 5)$ on the right-hand side of the zero axis in Figure 2.

The KS02 weighting scheme of Ω_{ks} appears to be the same as that of the partial moments of degree 1. Hence, Ω_{ks} of (6) is the ratio of the upper partial moment (UPM) of degree 1 to the analogous lower partial moment (LPM), Viole and Nawrocki (2016). In the R package called NNS, Viole (2021) has convenient functions called UPM and LPM to compute them, and hence

$$\Omega_{ks} = UPM(1, 0, r) / LPM(1, 0, r), \quad (7)$$

where r is a vector of returns, and where we have included the arguments (degree = 1, target = 0) of the R functions in the NNS package. Considering that these ratios are decades old, there is a need to compare the out-of-sample performance of Ω_{ks} from stock market return data using the NNS package with the three versions of Sharpe ratios in Equations (1), (3), and (4).

This subsection has explained the logic behind the cumbersome weighting scheme in KS02. The following subsection shows how the ECDF weighting is not really needed to achieve the intent in Bernardo and Ledoit (2000).

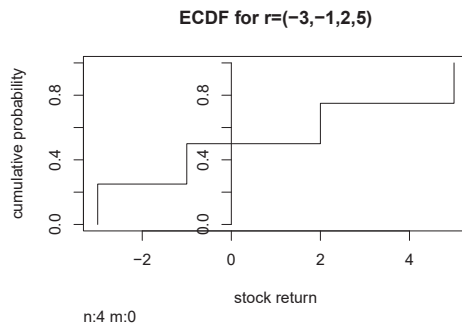


Figure 2. Empirical Cumulative Distribution Function for a Toy Example.

1.1.3. Recommended Computation of “Omega” for Stock Returns

Recall that we regard omega as one of many stock-picking algorithms. This subsection suggests a direct computation of Equation (5) for aggregate gain–aggregate loss ratios without using $F(r)$ and $(1 - F(r))$ weights. We divide the vector of returns r_t into positive (r_t^+) and negative (r_t^-) parts, proposing

$$\Omega_{sum} = \frac{\sum_t(r_t^+)}{\sum_t|r_t^-|}, \quad (8)$$

where the subscript ‘sum’ refers to summations in the formula. The numerator and denominator are both positive. For example, the stock A with returns $r_t^A = (-3, -1, 2, 5)$ has $\Omega_{sum} = ((2 + 5)/(3 + 1)) = 7/4$. The stock B with returns $r_t^B = (-4, 1, 2, 4)$ has the same $\Omega_{sum} = 7/4$. The computation of Ω_{sum} is seen to be easy and intuitive.

An alternative formulation is

$$\Omega_{avg} = \frac{\sum_t(r_t^+)/n^+}{\sum_t|r_t^-|/n^-}, \quad (9)$$

where n^+ denotes the number of positive returns in the data, n^- denotes the number of negative returns, and the subscript ‘avg’ refers to the averages in the formula. If the stock returns arose from an independent and identically distributed (IID) process, the probability of observing each return r_t is $1/n$ and averages equal mathematical expectations. Hence, it is tempting to prefer Ω_{avg} over Ω_{sum} . However, market returns are almost never IID. A stock’s return is intimately related to its own past, the returns of other stocks in its class, the socio-political conditions, etc. Since the joint density of all stocks is unknown, the probability of the sum of k returns ($\sum^k r_t$) is also unknown.

Can we use the averages in Ω_{avg} as an approximation by pretending that returns are IID? Let us compare the two Formulas (8) and (9). Observe that Ω_{avg} is larger than Ω_{sum} when n^- is large, and that Ω_{avg} is smaller when n^+ is large. The total gain–loss ratio is being multiplied by (n^-/n^+) . What are the implications of the relative sizes of (n^+, n^-) to the investor’s bottom line? We use two imaginary stocks, A and B, to argue that a larger (smaller) ratio (n^-/n^+) does not benefit (hurt) the bottom line of the investor and should be ignored.

Assume two sets of stock returns $r_t^A = (-3, -1, 2, 5)$ and $r_t^B = (-4, 1, 2, 4)$. They have the same $\Omega_{sum}^A = 1.75 = \Omega_{sum}^B$, though the $1.75 = \Omega_{avg}^A \gg \Omega_{avg}^B = 7/12 = 0.58333$. If one relies on Ω_{avg}^B , we have to conclude that the gain–pain ratio for stock A is over three times better than stock B. The aggregate gain (=7), aggregate loss (=4), and gain–loss ratio ($7/4 = 1.75$) to the investor is exactly the same for stocks A and B. The aggregate gain of 7 for stock B is spread over $n_B^+ = 3$ periods, while the aggregate loss (=4) is spread over only one period $n_B^- = 1$. The aggregate gain and loss for stock A is spread over two periods $n_A^+ = 2 = n_A^-$. The practically irrelevant (to the investor) sizes of (n^+, n^-) should not be allowed to contaminate the computation of the gain–loss ratio. We conclude this section by stating that there are sound reasons for rejecting Ω_{avg} of (9) in favor of the simpler Ω_{sum} of (8).

Table 3. Table of basic descriptive stats. ‘Sharpe’ is the ratio of mean to standard deviation (sd). Ω_s is Ω_{sum} , the sum of all positive returns divided by the sum of all negative returns. Number of non-missing or available sample size in the last column, ‘Av.N’. Part 1.

Ticker	Min	Q1	Median	Mean	Q3	Max	Sd	Sharpe	Ω_s	Av.N
aapl	−57.74	−4.66	2.53	2.40	9.81	45.38	12.20	0.20	1.7	473
amgn	−41.53	−3.62	1.59	2.17	6.36	45.88	9.94	0.22	1.9	473
amzn	−41.16	−4.73	2.47	3.61	9.74	126.38	16.56	0.22	2.0	324
apx	−32.09	−2.38	1.37	1.31	5.86	85.03	8.72	0.15	1.6	473
ba	−45.47	−3.82	1.45	1.18	6.99	45.93	8.83	0.13	1.4	473
cat	−35.91	−4.16	1.69	1.52	7.15	40.14	8.96	0.17	1.6	473
crm	−36.03	−4.61	1.75	2.39	8.98	40.26	11.08	0.22	1.8	239
cscv	−36.73	−4.01	1.70	2.19	8.27	38.92	10.48	0.21	1.8	411
cvx	−21.46	−2.57	1.09	1.15	4.76	26.97	6.47	0.18	1.6	473
dis	−28.64	−3.34	1.05	1.28	5.73	31.26	7.91	0.16	1.5	473
dow	−26.45	−3.37	1.24	0.86	6.30	25.48	9.41	0.09	1.3	62
gs	−27.73	−5.20	1.33	1.15	6.64	31.38	9.22	0.12	1.4	300
hd	−28.57	−3.41	1.71	1.91	7.06	30.33	8.14	0.23	1.8	473
hon	−38.19	−2.39	1.33	1.18	5.06	51.05	7.75	0.15	1.5	473
ibm	−24.86	−3.61	0.75	0.84	5.00	35.38	7.44	0.11	1.4	473

Table 4. Table of basic descriptive stats. Sharpe is the ratio of mean to standard deviation (sd). Ω_s is Ω_{sum} , the sum of all positive returns divided by the sum of all negative returns. Number of non-missing or available sample size in the last column, ‘Av.N’. Part 2.

Ticker	Min	Q1	Median	Mean	Q3	Max	Sd	Sharpe	Ω_s	Av.N
intc	−44.47	−4.39	1.25	1.52	7.05	48.81	10.67	0.14	1.5	473
jnj	−16.34	−2.23	1.25	1.22	4.43	19.29	5.59	0.22	1.8	473
jpm	−32.68	−3.86	1.22	1.32	6.36	33.75	9.17	0.14	1.5	473
ko	−19.33	−2.08	1.24	1.20	4.62	22.64	5.83	0.21	1.7	473
mcd	−25.67	−2.16	1.36	1.26	5.04	18.26	5.94	0.21	1.7	473
mmm	−27.83	−2.46	1.22	1.01	4.40	25.80	6.07	0.17	1.5	473
mrk	−26.62	−3.02	1.12	1.33	5.86	23.29	6.94	0.19	1.6	473
msft	−34.35	−3.56	2.25	2.36	6.80	51.55	9.45	0.25	2.0	458
nke	−37.50	−3.30	1.84	1.90	7.02	39.34	9.40	0.20	1.7	473
pg	−35.42	−1.77	1.20	1.20	4.92	24.69	5.57	0.22	1.8	473
trv	−53.47	−3.17	1.42	1.09	5.03	52.51	7.34	0.15	1.5	473
unh	−36.51	−3.21	2.50	2.18	7.39	40.70	9.57	0.23	1.9	473
v	−19.69	−2.47	2.27	1.58	5.25	16.83	6.14	0.26	1.9	194
vz	−20.48	−2.75	0.52	0.89	4.92	37.61	6.17	0.14	1.5	473
wmt	−27.06	−2.40	1.25	1.37	5.55	26.59	6.48	0.21	1.7	473

2. Unbiased Out-of-Sample Calculations

Most authors define their out-of-sample range from the last few periods of the data. Since any such out-of-sample time series is sensitive to the peculiar characteristics of the

last few periods, their calculations can be biased. For example, if the out-of-sample (**oos**) series coincides with the 2020 pandemic, the calculations will have a pessimistic bias. Vinod (2023) suggests removing the bias by “pandemic proofing” the calculations on (40% here) randomly chosen ‘oos’ data. Each j -th choice yields a ranking of stocks. Repeating the ranking J ($=50$, say) times, we compute their mean μ and standard deviation σ for the i -th stock-picking algorithm. We have $i = 1, \dots, 6$ here. Assuming no transaction costs, Vinod (2023) computes a zero-cost arbitrage, executing the following trades. One short sells (selling without first possessing) certain dollars worth of the worst stock in DJIA and buys the appropriate fraction of the best stock as determined by each method.

Our unbiased ‘oos’ strategy here does not seek a zero-cost arbitrage. Instead, we just compute distinct stock rankings by each method in-sample and randomly choose 40% observations for each j -th ‘oos’ realization. Finally, the average of the ranks over $j = 1, \dots, J$ random choices determines the ‘oos’ stock choice by that method.

3. Ranking 30 Stocks by Six Algorithms

We split our report into three tables, ranking the 30 DJIA stocks by two versions of six stock-picking algorithms. Each table has ten stocks at a time in alphabetical order of their ticker symbols. Tables 5–7 report two versions of the ranks produced by each algorithm for (a) **in-sample** ranks and (b) unbiased **out-of-sample** ranks.

1. **Sharpe-in/out:** Section 1.1.1 mentions the Sharpe ratio as a stock-picking algorithm.
2. **Omega-in/out:** Our computation is described in Section 1.1.3 and Equation (8) for Ω_{sum} .
3. **Decile-in/out:** It is generally agreed that the stock whose probability distribution of returns is more to the right-hand side is more desirable. One way to achieve this is to compare their deciles. The R package ‘generalCorr,’ offers a convenient function called `decileVote(.)`.
4. **Descr-in/out:** We compare the traditional descriptive stats of each stock’s data. Most stats are in the “the larger, the better” category and are obtained (+1) as weight. The standard deviation represents risk and obtains (−1) weight. This algorithm uses a weighted summary of these stats for stock-picking.
5. **Momen-in/out:** Moment values: The first four moments of a probability distribution provide information about centering, variability, skewness, and kurtosis. Our weights incorporate the prior knowledge that low variability and low kurtosis are desirable, while larger mean and skewness are desirable. A weighted summary is implemented in the R package ‘generalCorr,’ function called `momentVote(.)`.
6. **Exact-in/out:** This algorithm refers to the exact stochastic dominance mentioned in Section 3.1. The stochastic dominance (SD) of the first four orders is summarized in the R package ‘generalCorr.’ See the R function called `exactSd()`. The theoretical details are available in Vinod (2024), where iterated integrals of cumulative distribution functions are used. See the following Section 3.1 for a summary of general ideas behind the theory of stochastic dominance.

3.1. Ranking Stocks by Exact Stochastic Dominance

We apply some of the tools described in Vinod (2024) to the portfolio selection problem for the DJ30 data set used here. We use the exact computation of stochastic dominance using a worse imaginary stock ($x.ref$) than the worst-performing stock in DJ30.

We shall see that stochastic dominance needs a reference stock. Accordingly, we plan to compute the return for each of the 30 DJIA stocks with reference to the return in excess of the money-losing imaginary 31-st stock ($x.ref$). The lowest return over all included DJ30 data sets and over all 30 stocks is -86.14151 , where the negative sign suggests a loss. Let us choose ($x.ref$) return -87.38904 , which is a little smaller than -86.14151 , implying consistently the largest losses throughout the data period. Thus, all 30 stocks in our data always dominate ($x.ref$) throughout the period with varying dominance amounts.

The exact stochastic dominance computation invented by Vinod (2024) measures the dominance of each one of the thirty DJ30 stocks over the ($x.ref$) imaginary stock. The thirty

dominating amounts are comparable to each other and allow the ranking of the thirty stocks. The computation of dominating amounts depends on the order *k* of stochastic dominance (SD_k).

The first-order computation of the dominating amount depends on the exact area between two empirical cumulative distribution functions (ECDFs). It is customary to use iterated integrals for the higher-order computation of dominating areas since Levy (1973). The R package ‘generalCorr’ computes dominating areas from SD1 to SD4 and summarizes their rankings.

Tables 5–7 report the rank of stock tickers named in the column heading. The stocks ranked with low numbers (e.g., 1 to 8) are worth buying according to the algorithm implied by the row name. On the other hand, stocks ranked with high numbers (e.g., 23 to 30) are worth selling. Interestingly, stocks recommended for buying using in-sample data do not generally agree with the unbiased out-of-sample averages, even for the same algorithm.

Table 5. All criteria summary ranks of DJ30 stocks for in-sample and average over the randomized out-of-sample returns. Part 1.

Algorithm	aapl	amgn	amzn	axp	ba	cat	crm	csc	cvx	dis
Sharpe-in	15.0	6.0	7.0	22.0	27.0	18.0	8.0	12.0	17.0	20.0
Sharpe-out	29.0	23.0	24.0	15.0	20.0	22.0	13.0	8.0	9.0	21.0
Omega-in	15.0	5.0	2.0	18.0	27.0	19.0	9.0	11.0	17.0	22.0
Omega-out	29.0	22.0	24.0	15.0	21.0	23.0	14.0	7.0	11.0	20.0
Decile-in	9.0	15.0	4.5	7.0	22.5	13.0	10.5	6.0	28.0	26.0
Decile-out	25.0	21.5	24.0	4.0	11.0	6.5	19.0	8.0	26.5	23.0
Descr-in	14.3	16.0	13.0	12.0	17.0	14.6	14.1	14.9	17.6	18.0
Descr-out	21.9	17.4	16.6	14.6	13.6	15.9	13.0	14.9	16.6	17.7
Momen-in	7.0	3.0	1.0	10.0	24.0	11.5	4.0	6.0	20.0	15.0
Momen-out	19.0	10.0	7.0	14.0	23.0	21.5	3.0	1.0	17.0	20.0
Exact-in	8.0	10.0	5.0	18.0	25.0	14.0	4.0	6.0	26.0	19.0
Exact-out	29.0	18.0	5.0	13.0	21.0	22.0	2.0	6.0	19.0	23.0
AvgRank	21.0	13.0	7.0	11.0	27.0	16.0	6.0	3.0	22.0	24.0

Table 6. All criteria summary ranks of DJ30 stocks for in-sample and average over the randomized out-of-sample returns. Part 2.

Algorithm	dow	gs	hd	hon	ibm	intc	jnj	jpm	ko	mcd
Sharpe-in	30.0	28.0	3.0	21.0	29.0	26.0	5.0	24.0	13.0	10.0
Sharpe-out	30.0	27.0	7.0	6.0	28.0	26.0	4.0	10.0	1.0	17.0
Omega-in	30.0	28.0	6.0	21.0	29.0	25.0	8.0	24.0	10.0	13.0
Omega-out	30.0	27.0	8.0	5.0	28.0	25.0	4.0	12.0	1.0	17.0
Decile-in	24.5	27.0	8.0	14.0	30.0	21.0	18.0	24.5	16.0	12.0
Decile-out	16.5	16.5	10.0	5.0	30.0	28.0	12.0	2.0	9.0	26.5
Descr-in	19.6	18.7	12.7	15.4	19.6	17.7	14.4	18.1	14.7	14.0
Descr-out	18.3	16.7	13.9	14.6	20.1	19.3	13.6	13.9	11.6	17.0
Momen-in	30.0	25.0	8.0	16.0	28.0	11.5	18.0	17.0	26.0	23.0
Momen-out	30.0	27.0	5.0	11.0	24.0	12.5	18.0	6.0	9.0	29.0
Exact-in	1.0	3.0	11.0	24.0	30.0	13.0	21.0	17.0	22.0	20.0
Exact-out	1.0	4.0	9.0	12.0	30.0	20.0	17.0	10.0	11.0	27.0
AvgRank	28.0	26.0	4.0	12.0	30.0	25.0	10.0	14.0	8.0	23.0

Table 7. All criteria summary ranks of DJ30 stocks. Part 3.

Algorithm	mmm	mrk	msft	nke	pg	trv	unh	v	vz	wmt
Sharpe-in	19.0	16.0	2.0	14.0	9.0	23.0	4.0	1.0	25.0	11.0
Sharpe-out	19.0	16.0	14.0	25.0	3.0	12.0	11.0	2.0	5.0	18.0
Omega-in	20.0	16.0	1.0	14.0	7.0	23.0	4.0	3.0	26.0	12.0
Omega-out	19.0	16.0	13.0	26.0	3.0	9.0	10.0	2.0	6.0	18.0
Decile-in	22.5	19.0	3.0	2.0	17.0	20.0	1.0	4.5	29.0	10.5
Decile-out	18.0	20.0	29.0	14.0	13.0	6.5	1.0	3.0	15.0	21.5
Descr-in	18.3	16.3	11.9	13.7	16.7	16.9	11.7	11.7	17.1	14.3
Descr-out	17.7	16.9	14.7	18.1	10.7	14.7	11.6	11.1	11.6	17.0
Momen-in	29.0	19.0	2.0	9.0	21.0	22.0	5.0	13.0	27.0	14.0
Momen-out	28.0	21.5	2.0	26.0	12.5	15.0	4.0	16.0	8.0	25.0
Exact-in	28.0	16.0	7.0	12.0	23.0	27.0	9.0	2.0	29.0	15.0
Exact-out	28.0	24.0	7.0	25.0	14.0	16.0	8.0	3.0	15.0	26.0
AvgRank	29.0	20.0	5.0	15.0	9.0	18.0	2.0	1.0	19.0	17.0

The three tables, Tables 5–7, represent the main findings of our study of monthly data spanning over 39 years. The tables report alphabetically arranged ticker symbols in the column headings. Each table has ten stocks out of the thirty stocks of the DJIA. The bodies of the tables report rank by each of the twelve methods for six algorithms in two versions. For example, the first column of Table 5 has ranks of the Apple stock via the twelve methods. The first four ranks (15, 29, 15, 29) contain some repetitions, but the twelve methods rarely agree. If we respect the independent logic of all twelve algorithms, the average rank reported in the last (13-th) row may be thought of as a gist from the twelve ranks along the first twelve rows.

The buy or sell recommendations from our study needs some slicing and dicing of Tables 5–7. Accordingly, Table 8 reports abridged (case-sensitive single-character) names of the top eight stocks for buying (ranked 1 to 8) and the bottom eight for selling (ranked 23 to 30) by each of our six algorithms. Algorithm names are listed in Section 3. The algorithm names also appear as row names in Tables 5–7. We must abridge the row names to only three characters to save table space. Column names are lowercase versions of the stock ticker symbols. We are reporting twelve rows, two for each of the six criteria (in-sample and out-of-sample). The last row, named AvgRank, refers to the average rank from all 12 criteria listed above. Table 9 considers only the top and bottom two stocks for buying and selling.

Table 8. One character name of the top and bottom eight stocks for buying and selling by each of our six algorithms. Algorithm names from earlier tables are abridged to only three characters (suffix i = in-sample, o = out-of-sample). Column names are the ranks by the criterion named along the row.

algo	1	2	3	4	5	6	7	8	23	24	25	26	27	28	29	30
Shi	v	m	w	S	k	n	a	K	h	D	u	j	A	z	C	p
Sho	k	J	u	t	C	f	x	K	p	D	j	V	H	h	S	e
Omi	f	k	S	w	m	n	a	K	v	D	u	A	h	p	j	C
Omo	k	u	e	J	f	C	x	K	p	D	j	H	V	S	h	t
Dci	u	C	w	m	c	H	A	k	f	i	z	v	S	x	h	a
Dco	u	h	b	j	p	n	V	D	v	i	x	H	c	t	S	k
Dsi	u	v	f	x	h	z	n	m	e	I	d	J	M	g	D	i
Dso	p	v	k	u	V	C	b	j	A	d	M	n	D	I	i	a
Moi	z	x	c	I	v	w	d	H	A	D	C	u	S	a	h	n
Moo	S	A	H	I	p	x	t	v	C	D	u	h	J	z	V	k
Exi	D	A	h	n	I	c	w	K	g	i	C	z	S	f	a	u
Exo	D	J	k	H	x	p	V	t	v	i	g	z	S	f	u	h
avg	v	j	x	H	A	J	n	c	S	i	h	f	C	z	k	p

Table 9. Unabridged ticker symbols of the top two stocks for buying and the bottom two for selling by each of our six algorithms. Row names are algorithm names, as in earlier tables. Column names are ranks.

Criterion	1	2	29	30
Sharpe-in	v	mcd	crm	pg
Sharpe-out	ko	jpm	csc	cvx
Omega-in	msft	ko	jnj	crm
Omega-out	ko	unh	hd	trv
Decile-in	unh	crm	hd	aapl
Decile-out	unh	hd	csc	ko
Descr-in	unh	v	dow	ibm
Descr-out	pg	v	ibm	aapl
Momen-in	amzn	axp	hd	nke
Momen-out	csc	amgn	vz	ko
Exact-in	dow	amgn	aapl	unh
Exact-out	dow	jpm	unh	hd
AvgRank	v	jnj	ko	pg

4. Final Remarks

We describe the probability distribution of returns $f(r)$ and various stock-picking tools. We review Cheridito and Kromer’s **M**, **Q**, **S**, **D** properties. We reject their claim that the Sharpe ratio fails to satisfy monotonicity by showing the limitations in their proof. We also introduce new weights for the Omega ratio to address the distortions in the gain–pain ratio faced by investors.

This paper describes six stock-picking algorithms for long-term investment in the DJ30 stocks. Our implementations of omega (gain–pain ratio) and exact stochastic dominance appear to be new. We use monthly return data for the recent 39+ years to find that each algorithm leads to a distinct ranking. We report the ranks by each criterion, implying that rank 1 is the top stock worthy of buying and that rank 30 is the bottom stock worthy of selling. As we change the time periods (e.g., quarters, months, weeks, hours, etc.) included in the selected DJ30 data sets, the entire analysis will change, and our data-driven buy–sell recommendations are also expected to change. For example, Table 8 lists the top *eight* one-character abbreviations of ticker symbols to buy or sell. Table 9 lists the top *two* ticker symbols for stocks to buy and sell.

Our research shows that the ultimate choice of stock tickers to buy or sell in suitable quantities within one’s own budget is possible for anyone. Long-term investors need price data for long time intervals. It helps to compare many stock-picking algorithms along the lines shown here using a clear statement of the algorithms. The in-sample and unbiased out-of-sample ranks rarely coincide, even for the same algorithm. Averaging the ranks over six criteria using in-sample and an average of 50 randomized out-of-sample results suggests a choice of Visa and Johnson & Johnson for buying. The least desirable average ranks indicate that investors should consider selling Coca-Cola and Procter & Gamble shares.

Funding: This research received no external funding.

Data Availability Statement: Stock Prices of Dow Jones 30 stocks are publicly available. We use the R command `tseries::get.hist.quote` (“ticker symbol”, quote = “Adj”, start = “1985-01-01”, retclass = “zoo”, compression = “m”).

Acknowledgments: I thank Fred Viole and three anonymous referees for detailed and useful comments.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Online Computational Hints

Please use the link <https://faculty.fordham.edu/vinod/onlineDJ30jrfrm.docx> (accessed on 5 August 2024), where we provide online computational hints in the form of R code ready to copy and paste.

References

- Aumann, Robert J., and Roberto Serrano. 2008. An economic index of riskiness. *Journal of Political Economy* 116: 810–36. [CrossRef]
- Bernardo, Antonio E., and Olivier Ledoit. 2000. Gain, loss, and asset pricing. *Journal of Political Economy* 108: 144–72. [CrossRef]
- Cheridito, Patrick, and Eduard Kromer. 2013. Reward-risk ratios. *Journal of Investment Strategies* 3: 3–18. [CrossRef]
- Keating, Con, and William F. Shadwick. 2002. A Universal Performance Measure. Available online: <https://api.semanticscholar.org/CorpusID:157024871> (accessed on 5 August 2024).
- Levy, Haim. 1973. Stochastic dominance, efficiency criteria, and efficient portfolios: The multi-period case. *American Economic Review* 63: 986–94.
- Malkiel, Burton G. 2013. Asset management fees and the growth of finance. *Journal of Economic Perspectives* 27: 97–108. [CrossRef]
- Metel, Michael R., Traian A. Pirvu, and Julian Wong. 2017. Risk management under omega measure. *Risks* 5: 27. [CrossRef]
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Sharpe, William F. 1966. Mutual fund performance. *Journal of Business* 39 Pt 2: 119–38. [CrossRef]
- Vinod, Hrishikesh D. 2023. Pandemic-proofing Out-of-sample Portfolio Evaluations. *SSRN eLibrary*. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4542779 (accessed on 5 August 2024).
- Vinod, Hrishikesh D. 2024. Portfolio choice algorithms, including exact stochastic dominance. *Journal of Financial Stability* 70: 101196. [CrossRef]
- Vinod, Hrishikesh D., and Derrick Reagle. 2005. *Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments (Monograph)*. New York: Wiley.
- Viole, Fred. 2021. NNS: Nonlinear Nonparametric Statistics. R Package Version 0.8.3. Available online: <https://CRAN.R-project.org/package=NNS> (accessed on 5 August 2024).
- Viole, Fred, and David Nawrocki. 2016. LPM density functions for the computation of the sd efficient set. *Journal of Mathematical Finance* 6: 105–26. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

Action-Based Fiscal Consolidations and Economic Growth

Markus Brueckner

Research School of Economics, Australian National University, Acton, Canberra 0200, Australia;
markus.brueckner@anu.edu.au

Abstract: This paper tests the hypothesis that action-based fiscal consolidations have a negative effect on GDP growth. Using the IMF's dataset on action-based fiscal consolidations, instrumental variables' regressions show that action-based fiscal consolidations have a significant positive effect on GDP growth. The instrumental variables' regressions also show that action-based fiscal consolidations significantly increase investment and productivity. The findings presented in this paper thus strongly reject the hypothesis that action-based fiscal consolidations reduce growth. The paper argues that least squares estimates presented in previous literature suffer from negative reverse causality bias: GDP growth has a significant positive effect on both the likelihood and the magnitude of action-based fiscal consolidations. To uncover causal effects of action-based fiscal consolidations, researchers need to use an instrumental variables approach.

Keywords: fiscal consolidations; GDP growth; identification; narrative approach; simultaneous systems of equations

JEL Classification: E62

1. Introduction

This paper seeks to make an empirical contribution to the literature on the macroeconomic effects of fiscal consolidations. Specifically, the paper's main aim is to improve our understanding of what are the causal effects of so-called action-based fiscal consolidations. Theoretically, there is at least one compelling reason why action-based fiscal consolidations are not exogenous to contemporaneous GDP growth: automatic stabilizers. A decrease in GDP growth reduces tax revenues, and government expenditures automatically increase when GDP growth decreases (because decreases in GDP growth are generally associated with higher unemployment rates, which means that more people are seeking unemployment and welfare payments). An instrumental variables approach is thus needed for identifying the causal effects that action-based fiscal consolidations have on GDP growth.

This is the first paper to use an instrumental variables approach for estimating the causal effects that action-based fiscal consolidations have on GDP growth. The instrumental variables' estimates presented in this paper show that the response of GDP growth to action-based fiscal consolidations is positive and significantly different from zero: over a horizon of one year, a fiscal consolidation equal to 1 percent of GDP increases GDP by about 1.8 percent; over a horizon of two and three years, the cumulative GDP gain relative to the size of the cumulative fiscal consolidation shock is equal to about 1.5 and 1.4 percent, respectively.

The paper's finding is different from the mainstream finding in the existing literature, which is discussed in Section 2. The existing literature, to date, which has used least squares regressions, has found negative effects of fiscal consolidations on GDP growth. However, as argued in this paper, the findings in the existing literature are likely to be driven by negative reverse causality bias. Contemporaneous GDP growth has a significant negative effect on action-based fiscal consolidations. Least squares regressions, where GDP growth is the dependent variable and action-based fiscal consolidations is the explanatory variable,

Citation: Brueckner, Markus. 2024. Action-Based Fiscal Consolidations and Economic Growth. *Journal of Risk and Financial Management* 17: 194. <https://doi.org/10.3390/jrfm17050194>

Academic Editor: Thanasis Stengos

Received: 3 April 2024

Revised: 4 May 2024

Accepted: 6 May 2024

Published: 8 May 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

suffer from negative reverse causality bias. The instrumental variables' estimates that I report in this paper do not suffer from this negative reverse causality bias.

To shed light on the mechanisms through which action-based fiscal consolidations affect GDP growth, I examine the effects that action-based fiscal consolidations have on total factor productivity growth, investment, consumption, and net-exports. My instrumental variables' estimates show that action-based fiscal consolidations have a significant positive effect on total factor productivity growth, in addition to significantly boosting private investment and net-exports. I do not find a significant effect of action-based fiscal consolidations on private consumption. Action-based fiscal consolidations significantly reduce inflation and lead to a significant depreciation of the real exchange rate.

The paper's instrumental variables approach uses exclusion restrictions to identify a system of equations. At least since the 1980s (see, e.g., Hausman et al. (1987)), it has been known that, if valid, two exclusion restrictions are sufficient to identify a simultaneous system of two equations. In this paper, the two equations are: (1) where GDP growth contemporaneously affects action-based fiscal consolidations, and (2) where action-based fiscal consolidations contemporaneously affect GDP growth. Under the exclusion restriction of a zero covariance of the error terms in Equations (1) and (2), I only need valid external instruments for the first equation; then, the residual from that estimated equation is used as an instrument in the second equation to obtain a consistent estimate of the effect that action-based fiscal consolidations have on GDP growth.

I estimate the simultaneous system of two equations in two steps. In the first step, I estimate Equation (1) using temperature changes, GDP growth of trading partners, and the international commodity price index as excluded instruments for contemporaneous GDP growth. The instrumental variables' estimates show that GDP growth has, contemporaneously, a significant negative effect on the magnitude of fiscal consolidations. On average, a one percentage point increase in annual GDP growth decreases the magnitude of an action-based fiscal consolidation by about 0.08 percent of GDP in the same year. This is a large effect: the effect is equivalent to about one-quarter of the average action-based fiscal consolidation in the sample (in the sample, the average action-based fiscal consolidation is around 0.3 percent of GDP). I document that the estimated contemporaneous effect of GDP growth on fiscal consolidations is robust to using alternative instruments; that is, all three instruments jointly, or one instrument at a time, including additional control variables, such as lags of GDP growth and lags of fiscal consolidations, and excluding from the sample the five largest economies. In the second step, I use the residual variation in fiscal consolidations that is not due to GDP growth (i.e., the residual from the estimated Equation (1)) as an instrument for fiscal consolidations in Equation (2), where the dependent variable is GDP growth.

The remainder of the paper is organized as follows. Section 2 discusses related literature and clarifies the contribution and significance of this paper to the literature. Section 3 describes the data and estimation framework. Section 4 discusses the results. Section 5 shows that, by imposing exclusion restrictions, in a simultaneous system of equations, one can identify the effects that action-based fiscal consolidations have on GDP growth. Section 6 discusses instrumental variables' estimates of the simultaneous system of equations. Section 7 concludes the paper. The Supplementary Materials contain additional results.

2. Related Literature

For estimation of causal effects that fiscal consolidations have on GDP growth, a necessary condition is that fiscal consolidations are exogenous: contemporaneous GDP growth should have no effect on the magnitude or likelihood of a fiscal consolidation. This is the contemporaneous exogeneity condition (see, e.g., Stock and Watson (2018)). The literature is very much aware of the importance of this condition. Ramey (2016, 2019) provided an overview. One approach to identifying exogenous changes in fiscal policy, that has become increasingly popular, is the so-called narrative approach. According to the

narrative approach, the researcher reads policy documents—such as budgets and central bank reports—and then, based on what is written in those documents, classifies policy changes as exogenous if there is no indication in those documents that the policy change is in response to prospective economic conditions.

Does the absence of evidence—based on what is written in the policy documents—mean that the identified fiscal consolidations are exogenous to contemporaneous GDP growth? This paper provides an answer to that question. My results are for the largest dataset, in terms of country coverage, that exists: the IMF's dataset on action-based fiscal consolidations. This is a dataset where only those fiscal consolidations are recorded which, according to the narrative record, were primarily motivated by a desire to reduce the budget deficit. The first dataset on action-based fiscal consolidations was introduced in 2011 by DeVries et al. (2011). The dataset covered 17 advanced economies during the period 1978 to 2009. The data were extended to 14 countries in Latin America and the Caribbean during the period 1989 to 2016 by David and Leigh (2018). The IMF dataset covers both tax- and expenditure-based consolidations. Importantly, only those consolidations that were implemented as announced are included in the dataset.

The IMF dataset on fiscal consolidations has been widely used in the literature to estimate the effects of fiscal consolidations on GDP growth, either as the main variable (e.g., Guajardo et al. 2014; Jordà and Taylor 2016; Cloyne et al. 2020; Carrière-Swallow et al. 2021), as a robustness check (Pappa et al. 2015), or to construct fiscal plans (Alesina et al. 2015, 2018, 2019a, 2019b). In all those papers, the identifying assumption was that contemporaneous GDP growth has no effect on action-based fiscal consolidations.

Jordà and Taylor (2016) documented that year t GDP growth, as well as other macro-variables in year t , have significant effects on the year $t + 1$ likelihood of an action-based fiscal consolidation. The finding by Jordà and Taylor (2016), that action-based fiscal consolidations are predictable by GDP growth, does not imply that action-based fiscal consolidations are contemporaneously affected by GDP growth. Contemporaneous exogeneity is distinct from predictability. This point has already been made by Alesina et al. (2015, pp. 26–27), and it is worth reiterating it here. To the best of my knowledge, I am the first to examine whether action-based fiscal consolidations fulfill the contemporaneous exogeneity condition.

Providing an answer to the question of whether action-based fiscal consolidations are exogenous to contemporaneous GDP growth requires the use of an econometric model and country-specific variables that fulfill the following two conditions. First, the variables must be exogenous to GDP growth and fiscal consolidations. The second condition is that they should have a significant effect on GDP growth. Three candidate variables that likely fulfill these two conditions are: (i) year-to-year changes in temperature, (ii) the GDP growth rate of trading partners, and (iii) an international commodity price index. Among these three variables, the one that is clearly exogenous to the annual GDP growth rate of a country is the year-to-year change in the country-specific mean temperature. The second variable is exogenous to the GDP growth of a country, if that country's GDP is only a tiny fraction of the trading partners' GDP. The third variable is exogenous to the GDP growth of a country if that country is a price-taker on the international commodity market, i.e., the country exports or imports only a small fraction of the globally traded commodity.

Controlling for country and time fixed effects, my panel model estimates show that all three of these candidate variables individually have a significant effect on GDP growth, and they also individually have a significant effect on action-based fiscal consolidations. I am not the first to document that these three variables have a significant effect on annual GDP growth. There are numerous papers that have used these variables in various contexts, where one of the outcome variables is GDP growth.¹ I chose these three variables precisely because there exists a literature that has documented a significant effect of these variables on GDP growth.

A novel empirical result of this paper is that year-to-year changes in temperature, GDP growth of trading partners, and the international commodity price index have a significant contemporaneous effect on action-based fiscal consolidations. Importantly, the signs of the estimated effects are the same for all three of these variables; that is, each of these variables has a significant positive contemporaneous effect on annual GDP growth, as well as a significant negative contemporaneous effect on action-based fiscal consolidations. Taken together, these results suggest that it is highly unlikely that action-based fiscal consolidations are exogenous to contemporaneous GDP growth.

The reduced-form results are important for several reasons. First, they imply that the identifying assumption of contemporaneous exogeneity is not satisfied in those papers that have used the IMF's action-based fiscal consolidations variable on the right-hand side in econometric models, estimated by least squares, where the dependent variable is annual GDP growth. This is obviously an important result, as there is currently a lot of focus on fiscal policy and its effects on GDP growth. Second, the findings speak to the general question of whether a narrative approach (i.e., reading policy documents to look for absence of evidence) is suitable for identifying policy changes that are exogenous to contemporaneous GDP growth.² The answer to that question is no, the narrative approach is not suitable for identifying fiscal consolidations that are exogenous to contemporaneous GDP growth when the motive for the fiscal consolidation is budget-deficit reduction.

To be clear, there is immense value in the narrative approach. By reading policy documents, one can identify announcements and the main motive for the fiscal consolidation. The issue with the narrative approach that I am pointing out is that it is wrong to assume that tax or government spending changes, which are made by policymakers to reduce a budget deficit, are exogenous to contemporaneous GDP growth. There are other motives for fiscal policy changes: for example, wars fought overseas (see, e.g., Ramey (2011) or Ramey and Zubairy (2018)). My paper's results are specific to fiscal policy changes motivated by budget-deficit reduction. No inference should be made from the results in my paper regarding changes in fiscal policy, identified from the narrative record, that are due to motives other than budget-deficit reduction.

The narrative approach is not the only approach that exists in the literature for identifying causal effects of fiscal consolidations—there are approaches in the fiscal policy literature that use national accounts' data on government spending and tax revenues. One important paper is that of Blanchard and Perotti (2002). Blanchard and Perotti identified their system of equations by assuming that, at a quarterly frequency, government spending is contemporaneously unaffected by GDP growth. They used an outside estimate of the output elasticity of tax revenues to construct the residual variation in tax revenues that is not due to GDP, which they then used as an instrument for tax revenues. Another approach to identifying causal effects is to use sign restrictions (see, e.g., Mountford and Uhlig (2009)).

Romer and Romer's (2010) main critique of these alternative approaches was that there are likely to be variables, which are omitted from the econometric model, that have a direct contemporaneous effect on both GDP growth and fiscal variables. Romer and Romer advocated for the narrative approach. Their assumption was that, if the narrative record indicates that a fiscal policy change was mainly motivated by a desire to reduce an inherited budget deficit, then the fiscal policy change identified from the narrative record is exogenous to contemporaneous GDP growth. According to my estimates, this assumption can be rejected at the conventional significance levels.

Action-based fiscal consolidations are fundamentally different in nature to variations in government expenditures and tax revenues that arise due to other factors, such as, for example, automatic stabilizers. Fiscal consolidations are associated with a variety of issues that have been pointed out by the political economy literature (see, e.g., Alesina et al. (1998) for an early contribution). In that regard, it is not straightforward to relate the findings in this paper to the general question of how large are fiscal multipliers? Recent empirical work has uncovered a difference in the effects of fiscal expansions and fiscal consolidations (see,

e.g., Barnichon and Debortoli (2022)). The results in this paper are specific to action-based fiscal consolidations.

Two seminal theoretical macro-papers that are in line with the paper's empirical results are those of Giavazzi and Pagano (1990) and Blanchard (1990). Giavazzi and Pagano (1990) developed a model where an expenditure-based fiscal consolidation can have a positive effect on GDP growth. Blanchard (1990) showed that a positive effect on GDP growth is also possible for a tax-based fiscal consolidation.

An alternative, political economy explanation for why an action-based fiscal consolidation has a positive effect on GDP growth is as follows. Debt dynamics are determined by present-bias of policymakers (see, e.g., Yared (2019)). The larger the present-bias, the larger the distortions. If, throughout time, policymakers differ in their degree of present-bias, and their present-bias is private information, then a fiscal consolidation is a signal that the policymaker has a relatively low present-bias. Only those policymakers with a low present-bias will choose to reduce a budget deficit. The lower the present-bias, the higher the GDP growth.

3. Data and Estimation Framework for: Are Action-Based Fiscal Consolidations Exogenous?

3.1. Data

3.1.1. Action-Based Fiscal Consolidations

The data on action-based fiscal consolidation are from the International Monetary Fund (DeVries et al. 2011; David and Leigh 2018). The data were assembled by IMF economists following the narrative approach. In the working paper that accompanies the dataset, DeVries et al. (2011, pp. 3, 5) write:

"We examine policymakers' intentions and actions as described in contemporaneous policy documents, and identify measures motivated primarily by deficit reduction. . . Following Romer and Romer (2010), we use the contemporaneous estimates contained in these sources since retrospective estimates are rarely available."

According to Repec³, DeVries et al. (2011) have been cited in over 184 distinct research papers: more than 14 citations per year, on average, over a time span of 13 years since the release of the working paper. The number of citations suggests that many in the profession view the dataset as valuable. A database on discretionary fiscal policy aimed at consolidating the budget is extremely valuable. Such discretionary fiscal policies are different in nature to variations in the budget balance that arise due to automatic stabilizers. Discretionary fiscal policies aimed at consolidating the budget are actions taken by policymakers: tax-based consolidations are those actions by policymakers where tax rates are increased, while expenditure-based consolidations are those actions taken by policymakers that reduce government expenditures (broadly defined, i.e., purchases of goods and services, social transfers, and subsidies).

The IMF's database on action-based fiscal consolidations is the largest that exists. The data are publicly available and can be downloaded from the IMF's website. The dataset by DeVries et al. (2011) covers 17 advanced economies during 1978–2009, and the dataset by David and Leigh (2018) covers 14 Latin American and Caribbean economies during 1989–2016. In each dataset, there are three variables: the total value of the fiscal consolidation, tax-based fiscal consolidations, and expenditure-based fiscal consolidations. All three variables are expressed as a percent of a country's GDP. The data comprise only those fiscal consolidations that were implemented as announced. Fiscal consolidations that were announced and not implemented are not part of the dataset. The action-based fiscal consolidation data are annual.

3.1.2. Exogenous Variables

The data source for my baseline temperature variable was FAOSTAT (2021). As a robustness check, I also reported results for the temperature data of Dell et al. (2012). The baseline data on the trade-weighted GDP growth rate of trading partners and the international commodity price index were from Vegh and Vuletin (2015).⁴ As a robustness check, I also reported results for an international commodity price index that was constructed by the IMF economists Gruss and Kebhaj (2019). Specifically, I used the index that is based on time-invariant average GDP shares of commodity net-exports. I also reported the results for the net-barter terms of trade. Data on the net-barter terms of trade were from the World Development Indicators (World Bank 2021).

3.1.3. Descriptive Statistics

Appendix A, Table A1 shows descriptive statistics for an unbalanced panel of 31 countries during 1978–2016. The table reports descriptive statistics for the largest sample for which data are available on action-based fiscal consolidations. The descriptive statistics for the other variables were computed for a sample with observations that were either equal to or slightly smaller than those for which data are available on action-based fiscal consolidations. For Ecuador, there are no data in the Vegh and Vuletin (2015) dataset on GDP growth of trading partners and the international commodity price index. For Belgium, data on GDP growth of trading partners are available from 1998 onward, and the data from FAOSTAT on the temperature change are available from 2000 onward.

3.2. Estimation Framework

To examine exogeneity of the IMF's action-based fiscal consolidation variable, I used a panel model:

$$\text{FiscalConsolidation}_{it} = a_i + b_t + \alpha Z_{it} + u_{it} \quad (1)$$

where *FiscalConsolidation* is the IMF's action-based fiscal consolidation variable in year *t* and country *i*. In the sections that follow, where I discuss estimates of the above equation, I have used fiscal consolidations as a short-hand for the IMF's action-based fiscal consolidation variable. The model includes country and year fixed effects, denoted by a_i and b_t , respectively. *Z* are variables that are exogenous to fiscal consolidations. In Equation (1), rejecting the null hypothesis that $\alpha = 0$ means that the IMF's action-based fiscal consolidations are unlikely to be exogenous.

The choice of variables for *Z* was motivated by the literature, which has documented a significant effect of these variables on GDP growth. The three variables were: the year $t - 1$ to t change in the mean annual temperature of country *i*, the year *t* growth rate of the GDP of trading partners of country *i*, and the year *t* international commodity price index for country *i*. To obtain an estimate of the average effect of these variables on GDP growth for the sample at hand, I estimated Equation (2) for the same number of observations as Equation (1):

$$\text{GDPGrowth}_{it} = c_i + b_t + \beta * Z_{it} + e_{it} \quad (2)$$

Conditional on rejecting the null $\alpha = 0$ in Equation (1), if in Equation (2) one also rejects that $\beta = 0$, then it is unlikely that—in models where the dependent variable is GDP growth and the right-hand-side variable is fiscal consolidation—fiscal consolidations fulfill the contemporaneous exogeneity condition. The argument is straightforward. First, it is implausible to think that any of the variables in *Z* affect GDP growth because of their effect on fiscal consolidations. Second, it is plausible that these variables affect consolidations because of their effect on GDP growth: temperature changes affect GDP growth because of their effect on changes in agricultural productivity, GDP growth of trading partners affects the GDP growth of a country because it affects the demand for that country's exports, and the international commodity price index affects GDP growth of a country because it affects the country's terms of trade, which in turn affects the country's value of net-exports.⁵

4. Empirical Results for: Are Action-Based Fiscal Consolidations Exogenous?

4.1. Contemporaneous Effects of Temperature Changes

Table 1 shows that the year $t - 1$ to t change of annual mean temperatures has significant contemporaneous effects on both the magnitude of fiscal consolidations and on GDP growth. For the estimates shown in Table 1, the temperature data were from FAOSTAT (2021).

Table 1. Contemporaneous effects of temperature changes on action-based fiscal consolidations and GDP growth.

	Fiscal Consolidation	Fiscal Consolidation	GDP Growth	GDP Growth
	(1)	(2)	(3)	(4)
	Whole Sample	Excluding Large Economies	Whole Sample	Excluding Large Economies
Temperature Change, t	−0.07 ** (0.04)	−0.10 ** (0.04)	0.71 *** (0.2)	0.75 *** (0.22)
Country Fixed Effects	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes
Observations	986	805	986	805
Countries	31	26	31	26

Note: The dependent variable in columns (1) and (2) is *Fiscal Consolidation* in year t ; in columns (3) and (4), the dependent variable is *GDP growth* in year t . The method of estimation is least squares. Columns (1) and (3) show estimates for the whole sample. Columns (2) and (4) show estimates for the sub-sample that excludes France, Germany, Japan, the United Kingdom, and the United States. Huber robust standard errors (shown in parentheses) are clustered at the country level. * Significantly different from zero at the 10 percent level, ** 5 percent level, and *** 1 percent level.

The first two columns of Table 1 show that temperature changes have a significant negative contemporaneous effect on the magnitude of fiscal consolidations. Specifically, in column (1) of Table 1, one can see that the estimated coefficient on temperature changes is around -0.07 . The estimated coefficient has a standard error of around 0.04 . Thus, one can reject the hypothesis that the coefficient is equal to zero at the 10 percent significance level (p -value 0.06). Column (2) reports estimates for the sub-sample that excludes the five largest economies. One can see that in this sub-sample, the contemporaneous effect of temperature changes on fiscal consolidations is somewhat larger in absolute. That is, in column (2), the coefficient on temperature changes is around -0.10 and has a standard error of 0.04 . Therefore, one can reject the hypothesis that this coefficient is equal to zero at the 5 percent level (p -value 0.02). Quantitatively, the estimates reported in columns (1) and (2) suggest that a one standard deviation increase in the year $t - 1$ to t change in temperature decreases the magnitude of a fiscal consolidation in year t by around 0.04 to 0.06 percent of GDP.

Columns (3) and (4) of Table 1 show that temperature changes have a significant positive contemporaneous effect on GDP growth in the sample in which temperature changes have a significant negative contemporaneous effect on fiscal consolidations. The country-year observations in columns (3) and (4) are the same as in columns (1) and (2), respectively. In column (3) of Table 1, one can see that the estimated coefficient on temperature changes is around 0.71 . The estimated coefficient has a standard error of around 0.20 . Thus, one can reject the hypothesis that the coefficient is equal to zero at the 1 percent significance level (p -value 0.00). Column (4) reports the estimated coefficient on temperatures for the sub-sample that excludes the five largest economies. One can see that in this sub-sample, the contemporaneous effect of temperature changes on GDP growth is somewhat larger in absolute. That is, in column (4), the coefficient on temperature changes is 0.75 and has a standard error of 0.22 . Hence, one can reject the hypothesis that this coefficient is equal to zero at the 1 percent level (p -value 0.00). Quantitatively, the estimates reported in columns (3) and (4) suggest that a one standard deviation increase in the year $t - 1$ to t change in temperature increases GDP growth in year t by around 0.4 percentage points.

4.2. Contemporaneous Effects of GDP Growth of Trading Partners

Table 2 shows that the GDP growth rate of trading partners has significant contemporaneous effects on both the magnitude of fiscal consolidations and on GDP growth. The first two columns of Table 2 show that the GDP growth rate of trading partners has a significant negative contemporaneous effect on the magnitude of fiscal consolidations. Specifically, in column (1) of Table 2, one can see that the estimated coefficient on GDP growth of trading partners is around -0.25 . The estimated coefficient has a standard error of around 0.08 . Therefore, one can reject the hypothesis that the coefficient is equal to zero at the 1 percent significance level (p -value 0.01). Column (2) shows that the estimated coefficient on GDP growth of trading partners is similar for the sub-sample that excludes the five largest economies. One can see in column (2) that the coefficient on GDP growth of trading partners is around -0.25 and has a standard error of 0.09 . Thus, one can reject the hypothesis that this estimated coefficient is equal to zero at the 1 percent level (p -value 0.01). Quantitatively, the estimates reported in columns (1) and (2) suggest that a one standard deviation increase in GDP growth of trading partners in year t decreases the magnitude of a fiscal consolidation in year t by around 0.13 percent of GDP.

Table 2. Contemporaneous effects of GDP growth of trading partners on action-based fiscal consolidations and GDP growth.

	Fiscal Consolidation	Fiscal Consolidation	GDP Growth	GDP Growth
	(1)	(2)	(3)	(4)
	Whole Sample	Excluding Large Economies	Whole Sample	Excluding Large Economies
GDP Growth of Trading Partners, t	-0.25^{***}	-0.25^{***}	2.65^{***}	2.61^{***}
	(0.08)	(0.09)	(0.45)	(0.46)
Country Fixed Effects	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes
Observations	913	733	913	733
Countries	30	25	30	25

Note: The dependent variable in columns (1) and (2) is *Fiscal Consolidation* in year t ; in columns (3) and (4), the dependent variable is *GDP growth* in year t . The method of estimation is least squares. Columns (1) and (3) show estimates for the whole sample. Columns (2) and (4) show estimates for the sub-sample that excludes France, Germany, Japan, the United Kingdom, and the United States. Huber robust standard errors (shown in parentheses) are clustered at the country level. * Significantly different from zero at the 10 percent level, ** 5 percent level, and *** 1 percent level.

Columns (3) and (4) of Table 2 show that GDP growth of trading partners has a significant positive contemporaneous effect on GDP growth in the sample in which GDP growth of trading partners has a significant negative contemporaneous effect on fiscal consolidations. The country–year observations in columns (3) and (4) are the same as those in columns (1) and (2), respectively. In column (3) of Table 2, one can see that the estimated coefficient on GDP growth of trading partners is around 2.65 . The estimated coefficient has a standard error of around 0.45 . Thus, one can reject the hypothesis that the coefficient is equal to zero at the 1 percent significance level (p -value 0.00). Column (4) reports the estimated coefficient on GDP growth of trading partners for the sub-sample that excludes the five largest economies. The estimated coefficient is 2.61 and has a standard error of 0.46 . Therefore, one can reject the hypothesis that this coefficient is equal to zero at the 1 percent level (p -value 0.00). Quantitatively, the estimates reported in columns (3) and (4) suggest that a one standard deviation increase in the GDP growth of trading partners in year t increases GDP growth in year t by around 1.4 percentage points.

4.3. Contemporaneous Effects of an International Commodity Price Index

Table 3 shows that a country-specific international commodity price index has significant contemporaneous effects on both the magnitude of fiscal consolidations and on GDP growth. For the estimates shown in Table 3, the country-specific international commodity price index is from Vegh and Vuletin (2015).

Table 3. Contemporaneous effects of an international commodity price index on action-based fiscal consolidations and GDP growth.

	Fiscal Consolidation	Fiscal Consolidation	GDP Growth	GDP Growth
	(1)	(2)	(3)	(4)
	Whole Sample	Excluding Large Economies	Whole Sample	Excluding Large Economies
Commodity Price Index, <i>t</i>	−0.02 ** (0.01)	−0.02 ** (0.01)	0.20 *** (0.06)	0.18 *** (0.07)
Country Fixed Effects	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes
Observations	908	728	908	728
Countries	30	25	30	25

Note: The dependent variable in columns (1) and (2) is *Fiscal Consolidation* in year *t*; in columns (3) and (4), the dependent variable is *GDP growth* in year *t*. The method of estimation is least squares. Columns (1) and (3) show estimates for the whole sample. Columns (2) and (4) show estimates for the sub-sample that excludes France, Germany, Japan, the United Kingdom, and the United States. Huber robust standard errors (shown in parentheses) are clustered at the country level. * Significantly different from zero at the 10 percent level, ** 5 percent level, and *** 1 percent level.

The first two columns of Table 3 show that the commodity price index has a significant negative contemporaneous effect on the magnitude of fiscal consolidations. Specifically, in column (1) of Table 1, the estimated coefficient on the commodity price index is −0.017. The estimated coefficient has a standard error of 0.007. Thus, one can reject the hypothesis that this coefficient is equal to zero at the 5 percent significance level (*p*-value 0.03). Column (2) reports estimates for the sub-sample that excludes the five largest economies. In this sub-sample, the contemporaneous effect of the commodity price index is slightly larger in absolute. That is, in column (2), the coefficient on the commodity price index is around −0.018 and has a standard error of 0.008. Therefore, one can reject the hypothesis that this estimated coefficient is equal to zero at the 5 percent level (*p*-value 0.04). Quantitatively, the estimates reported in columns (1) and (2) suggest that a one standard deviation increase in the commodity price index in year *t* decreases the magnitude of a fiscal consolidation in year *t* by around 0.10 percent of GDP.

Columns (3) and (4) of Table 3 show that the commodity price index has a significant positive contemporaneous effect on GDP growth in the sample in which the commodity price index has a significant negative contemporaneous effect on fiscal consolidations. The number of observations in columns (3) and (4) are the same as in columns (1) and (2), respectively. In column (3) of Table 3, one can see that the estimated coefficient on the commodity price index is around 0.20. The estimated coefficient has a standard error of around 0.06. Thus, one can reject the hypothesis that the estimated coefficient is equal to zero at the 1 percent significance level (*p*-value 0.01). Column (4) reports the estimated coefficient for the sub-sample that excludes the five largest economies. In this sub-sample, the coefficient on the commodity price index is 0.18 and has a standard error of 0.06. Therefore, one can reject the hypothesis that this coefficient is equal to zero at the 1 percent level (*p*-value 0.01). Quantitatively, the estimates reported in columns (3) and (4) suggest that a one standard deviation increase in the commodity price index in year *t* increases GDP growth in year *t* by around 1 percentage point.

5. Estimation Framework: Identification of a Simultaneous System of Two Equations

Consider the simplest possible simultaneous system of two equations:

- (1) $FiscalConsolidation = \alpha GDPGrowth + u,$
- (2) $GDPGrowth = \beta * FiscalConsolidation + e,$

where I have dropped subscripts to keep the notation as simple as possible. In the spirit of the Frisch–Waugh–Lovell theorem, one can think of the above variables as the residuals obtained from separate regressions of GDP growth and fiscal consolidations on a set of control variables, such as country and time fixed effects, past GDP growth, and past fiscal consolidations.

Under the assumption that $cov(u,e) = 0$, the least squares estimate of β in Equation (2) is:

$$(I) \quad \beta^{LS} = \beta + cov(FiscalConsolidation, u) = \beta + \alpha(1 - \alpha\beta)^{-1}\sigma^2(e)/\sigma^2(F),$$

where $\sigma^2(e)$ is the variance of the error term e and $\sigma^2(F)$ is the variance of fiscal consolidations. Hence, only if α is equal to zero does the least squares estimation of Equation (2) yield a consistent estimate of β .

To identify the simultaneous system of two equations, I need at least one variable, denoted here by Z , that satisfies the following conditions: (i) Z is exogenous to GDP growth and fiscal consolidations, (ii) Z affects GDP growth, and (iii) Z affects fiscal consolidations through its effect on GDP growth:

$$(2') \quad GDPGrowth = \beta * FiscalConsolidation + \phi Z + e'$$

Under conditions (i)–(iii), Z can be used as an instrument for GDP growth to obtain a consistent estimate of α in Equation (1).

Note that from Equations (1) and (2'), it follows that the reduced-form effect of Z on fiscal consolidations is $\alpha\phi$. Conditional on GDP growth, the effect of Z on fiscal consolidations is 0. That is, Z only affects fiscal consolidations through GDP growth. There is no direct effect of Z on fiscal consolidations.

Based on previous literature, I have used three variables as candidates for Z : temperature changes, the trade-weighted GDP growth rate of trading partners, and an international commodity price index. In the context of estimating the effects of annual GDP growth on fiscal variables, there are a number of papers that have used these variables as instruments for GDP growth. However, in none of those papers was the outcome variable the IMF's action-based fiscal consolidations variable. Brueckner (2012) used an international commodity price index as an instrument for GDP growth to estimate the elasticity response of tax revenues to GDP growth for a panel of 33 Sub-Saharan African countries during the period 1980–2000. Brueckner et al. (2012) used an international oil price index as an instrument for GDP growth to estimate the contemporaneous elasticity response of government spending to GDP growth. Vegh and Vuletin (2015) used the GDP growth rate of trading partners and an international commodity price index as instruments for GDP growth when estimating the contemporaneous effect of GDP growth on tax rates, and their panel consisted of 62 countries during the period 1960–2013. In all the above papers, the assumption was that the commodity price index and GDP growth of trading partners only affect fiscal variables through their effect on GDP growth.

Once an estimate of α is obtained (from an instrumental variables' regression, where GDP growth is instrumented by Z), the next step is to construct the residual variation in fiscal consolidations that is not due to GDP growth, i.e., $u_{res} = FiscalConsolidation - \alpha^{hat, IV} GDPGrowth$. Then, u_{res} is used as an instrument for fiscal consolidations in Equation (2). This yields the following instrumental variables' estimator:

$$(II) \quad \begin{aligned} \beta^{IV} &= cov(u_{res}, GDPGrowth) / cov(u_{res}, FiscalConsolidation) \\ &= \beta + cov(u_{res}, e) / cov(u_{res}, FiscalConsolidation) \\ &= \beta + cov(u, e) / cov(u, FiscalConsolidation) \\ &= \beta \end{aligned}$$

where line three demonstrates that $\alpha^{\text{hat,IV}} = \alpha$, and hence $u_{\text{res}} = u$. The last line follows from the assumption that $\text{cov}(u, e) = 0$. This is the same assumption that I made to derive the OLS estimator in Equation (I).

Note that this instrumental variables strategy to identify a simultaneous system of two equations yields consistent estimates if one has at least one valid instrument, Z . One cannot identify the simultaneous system of equations by using least squares estimation of Equation (1), computing the residual, and then using that residual as an instrument in Equation (2). The reason why this will yield inconsistent estimates of β is that $\alpha^{\text{hat,LS}} \neq \alpha$, from which it follows that $u'_{\text{res}} = \text{FiscalConsolidation} - \alpha^{\text{hat,LS}} \text{GDPGrowth} \neq u$, and thus $\text{cov}(u'_{\text{res}}, e) \neq 0$.

6. Instrumental Variables' Estimates of the Simultaneous System of Equations

6.1. The Contemporaneous Effect of GDP Growth on Fiscal Consolidations

6.1.1. Overidentified Model

Table 4 reports two-stage least squares estimates of the contemporaneous effect that GDP growth has on the magnitude of fiscal consolidations. The models in Table 4 are overidentified: there is one endogenous variable—GDP growth, and three instruments, namely, temperature changes, the GDP growth rate of trading partners, and the international commodity price index. The top panel of Table 4 reports the second-stage estimates (the effect that GDP growth in year t has on the magnitude of fiscal consolidations in year t), while the bottom panel reports the first-stage estimates (of the effect that the instruments have on GDP growth).

The main message of the IV estimates reported in Table 4 is that contemporaneous GDP growth has a significant negative effect on the magnitude of fiscal consolidations. Column (1) of Table 4 reports estimates for the largest sample considering the available data. In that column, the estimated coefficient on year t GDP growth is -0.07 and has a standard error of 0.02 . Thus, one can reject the hypothesis that the coefficient is equal to zero at the 1 percent significance level (p -value 0.003). The Anderson–Rubin test rejects that the coefficient on GDP growth in column (1) is equal to zero at the 5 percent level (p -value 0.031). Quantitatively, the estimated coefficient of -0.07 suggests that, on average, a one percentage point increase in annual GDP growth in year t decreases the magnitude of a fiscal consolidation by around 0.07 percent of GDP.

Standard test diagnostics suggest that the 2SLS estimates are based on instruments that are relevant and valid. The Kleibergen–Paap F -stat and Cragg–Donald F -stat for the hypothesis that the first-stage effects of the three instruments are jointly equal to zero is 17.0 and 19.7 , respectively. The p -value of the Hansen J test, showing that the three instruments are jointly uncorrelated with the second-stage error term, is 0.99 .

Column (2) of Table 4 reports two-stage least squares estimates for the sub-sample that excludes the five largest economies. Excluding the five largest economies led to a slightly larger negative effect of GDP growth on fiscal consolidations. In column (2), the coefficient on GDP growth is -0.08 and has a standard error of 0.03 . Therefore, one can reject the hypothesis that the coefficient in column (2) is equal to zero at the 1 percent significance level (p -value 0.002). The Anderson–Rubin test rejects that the coefficient on GDP growth in column (2) is equal to zero at the 5 percent level (p -value 0.037).

Contemporaneous GDP growth has a significant negative effect on both tax- and expenditure-based fiscal consolidations. This can be seen from the estimates in columns (3)–(6) of Table 4. Contemporaneous GDP growth has a somewhat larger negative effect on tax-based consolidations than on expenditure-based consolidations. A one percentage point increase in year t GDP growth reduces the magnitude of a tax-based fiscal consolidation by around 0.05 percent of GDP. That is, for the purpose of consolidating the budget, tax rates must increase when GDP growth decreases. For an expenditure-based consolidation, this effect amounts to around 0.03 percent of GDP. That is, for the purpose of consolidating the budget, discretionary government expenditures must decrease when GDP growth decreases.

Table 4. Contemporaneous effects of GDP growth on action-based fiscal consolidations.

	Fiscal Consolidation (Tax and Expenditure)	Fiscal Consolidation (Tax and Expenditure)	Fiscal Consolidation (Tax)	Fiscal Consolidation (Tax)	Fiscal Consolidation (Tax)	Fiscal Consolidation (Expenditure)	Fiscal Consolidation (Expenditure)
	(1)	(2)	(3)	(4)	(5)	(6)	(6)
	Whole Sample	Excluding Large Economies	Whole Sample	Excluding Large Economies	Whole Sample	Excluding Large Economies	Excluding Large Economies
GDP Growth, <i>t</i>	−0.07 ** (0.02)	−0.08 *** (0.03)	−0.05 ** (0.02)	−0.05 ** (0.02)	−0.03 ** (0.01)	−0.03 * (0.02)	−0.03 * (0.02)
Cragg–Donald F-Stat	19.7	13.9	19.7	13.9	19.7	13.9	13.9
Kleibergen–Paap F-Stat	17	13.7	17	13.7	17	13.7	13.7
Hansen <i>J</i> , <i>p</i> -value	0.99	0.75	0.87	0.79	0.63	0.37	0.37
	First-Stage Estimates for GDP Growth, <i>t</i>						
GDP Growth of Trading Partners, <i>t</i>	2.21 *** (0.51)	2.13 *** (0.52)	2.21 *** (0.51)	2.13 *** (0.52)	2.21 *** (0.51)	2.13 *** (0.52)	2.13 *** (0.52)
Commodity Price Index, <i>t</i>	0.17 ** (0.07)	0.14 * (0.07)	0.17 ** (0.07)	0.14 * (0.07)	0.17 ** (0.07)	0.14 * (0.07)	0.14 * (0.07)
Temperature Change, <i>t</i>	0.71 *** (0.22)	0.71 *** (0.26)	0.71 *** (0.22)	0.71 *** (0.26)	0.71 *** (0.22)	0.71 *** (0.26)	0.71 *** (0.26)
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	850	670	850	670	850	670	670
Countries	29	24	29	24	29	24	24

Note: The method of estimation is two-stage least squares. Huber robust standard errors (shown in parentheses) are clustered at the country level. Columns (1), (3), and (5) show estimates for the whole sample. Columns (2), (4), and (6) show estimates for the sub-sample that excludes France, Germany, Japan, the United Kingdom, and the United States. * Significantly different from zero at the 10 percent level, ** 5 percent level, and *** 1 percent level.

6.1.2. Just-Identified Model

Instrumental variables’ estimation provides a local average treatment effect. The question arises, then, whether the finding in the previous section, which was obtained from an overidentified model, is due to the nature of any one of the three instruments. To provide an answer to that question, Table 5 reports estimates of just-identified models. In these models, the dependent variable is the magnitude of fiscal consolidations in year t , and the endogenous right-hand-side variable is GDP growth in year t .

Table 5. Contemporaneous effects of GDP growth on action-based fiscal consolidations (two-stage least squares with one instrument for GDP growth).

	Fiscal Consolidation					
	(1)	(2)	(3)	(4)	(5)	(6)
	Whole Sample	Excluding Large Economies	Whole Sample	Excluding Large Economies	Whole Sample	Excluding Large Economies
GDP Growth, t	−0.09 *** (0.03)	−0.10 *** (0.03)	−0.08 ** (0.04)	−0.10 ** (0.05)	−0.10 * (0.05)	−0.14 ** (0.06)
Cragg–Donald F-Stat	48.8	13.7	27.9	13.7	8.7	7
Kleibergen–Paap F-Stat	33.65	13.9	9.1	13.9	12.9	11.1
First-Stage Estimates for GDP Growth, t						
GDP Growth of Trading Partners, t	2.65 *** (0.45)	2.61 *** (0.45)				
Commodity Price Index, t			0.20 *** (0.07)	0.18 *** (0.07)		
Temperature Change, t					0.71 *** (0.19)	0.75 *** (0.22)
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	913	733	908	728	990	670
Countries	30	24	30	25	31	24

Note: The dependent variable is *Fiscal Consolidation* in year t . The method of estimation is two-stage least squares. Huber robust standard errors (shown in parentheses) are clustered at the country level. Columns (1), (3), and (5) show estimates for the whole sample. Columns (2), (4), and (6) show estimates for the sub-sample that excludes France, Germany, Japan, the United Kingdom, and the United States. * Significantly different from zero at the 10 percent level, ** 5 percent level, and *** 1 percent level.

The estimated models in Table 5 are just-identified models: there is one instrument for one endogenous variable, GDP growth. Specifically, in columns (1) and (2) of Table 5, the instrument for GDP growth in year t is the GDP growth rate of trading partners in year t . In columns (3) and (4), the instrument for year t GDP growth is the year t international commodity price index. In columns (5) and (6), the instrument for year t GDP growth is the year $t - 1$ to t change in temperature change. Columns (1), (3), and (5) show results for the largest sample, and columns (2), (4), and (6) show results for the sample that excludes the five largest economies. The top panel of Table 5 reports the estimated second-stage coefficient on GDP growth. The bottom panel reports the first-stage effect that the instrument has on GDP growth.

From the top panel of Table 5, one can see that the second-stage estimates on GDP growth in the just-identified 2SLS regressions are similar—both quantitatively and statistically—across the three different instruments. In column (1), where the excluded instrument is the GDP growth rate of trading partners, the estimated coefficient on year t GDP growth is -0.09 and has a standard error of 0.03 . In column (3), where the excluded instrument is the international commodity price index, the estimated coefficient on year t GDP growth is -0.08 and has a standard error of 0.04 . When the instrument is the temperature change (see column (5)), the estimated coefficient on year t GDP growth is -0.10 and has a standard error of 0.05 . Repeating the same regressions for the sub-sample that excludes the five largest economies yields estimated coefficients on GDP growth of -0.10 , -0.10 , and -0.14 , respectively (see columns (2), (4), and (6)).

The similarity of the estimated coefficients on GDP growth in the just-identified models, as shown in Table 5, suggests that the estimates of the overidentified models,

as shown in Table 4, are not driven by a particular instrument. If it were the case that a negative effect of GDP growth on fiscal consolidations is due to the nature of any one of the three instruments, then one would see very different estimates on GDP growth in just-identified models depending on what instrument was used. Table 5 demonstrates that this is not the case.

6.1.3. Tests for Direct Effects of the Instruments on Fiscal Consolidations

There are exclusion restrictions underlying the instrumental variables’ regressions, namely, that the instruments only affect fiscal consolidations through GDP growth. Tables 1–3 showed that there are significant negative reduced-form effects of each instrument on fiscal consolidations. The exclusion restriction of no direct effect means that conditional on GDP growth, temperature changes, GDP growth of trading partners, and the international commodity price index have no effect on fiscal consolidations. Testing this requires at least one instrument for GDP growth that is relevant (i.e., has a significant effect on GDP growth), exogenous to GDP growth and fiscal consolidations, and only affects fiscal consolidations through GDP growth. Tests of direct effects of the instruments can only be made conditional on the assumption that a sub-set of the instruments are relevant and valid.

Table 6 shows that GDP growth of trading partners, the international commodity price index, and temperature changes have no significant direct effects on fiscal consolidations. This is the case for the largest sample (columns (1)–(3)) and the sub-sample that excludes the five largest economies (columns (4)–(6)). Conditional on GDP growth, the estimated effects that GDP growth of trading partners, the international commodity price index, and temperature changes have on fiscal consolidations are quantitatively small and are not significantly different from zero at the conventional significance levels.

Table 6. Contemporaneous effects of GDP growth on action-based fiscal consolidations (examination of the exclusion restrictions).

	Fiscal Consolidation					
	(1)	(2)	(3)	(4)	(5)	(6)
	Whole Sample			Excluding Large Economies		
GDP Growth, <i>t</i>	−0.08 *** (0.02)	−0.09 * (0.05)	−0.07 ** (0.03)	−0.08 ** (0.02)	−0.11 ** (0.06)	−0.07 ** (0.03)
Temperature Change, <i>t</i>	0.01 (0.04)			−0.01 (0.04)		
GDP Growth of Trading Partners, <i>t</i>		0.04 (0.17)			0.1 (0.2)	
Commodity Price Index, <i>t</i>			−0.00 (0.01)			−0.01 (0.01)
Hansen J, <i>p</i> -value	0.73	0.68	0.98	0.61	0.86	0.61
Cragg–Donald F-Stat	29.3	11.1	21.7	20.9	7.3	17
Kleibergen–Paap F-Stat	16.1	10.4	23	14	7.4	21
	First-Stage Estimates for GDP Growth					
Temperature Change, <i>t</i>		0.73 *** (0.21)	0.73 *** (0.21)		0.74 *** (0.25)	0.74 *** (0.25)
GDP Growth of Trading Partners, <i>t</i>	2.28 *** (0.47)		2.28 *** (0.47)	2.25 *** (0.49)		2.25 *** (0.49)
Commodity Price Index, <i>t</i>	0.16 *** (0.07)	0.16 ** (0.07)		0.13 * (0.07)	0.13 * (0.07)	
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	882	882	882	702	702	702
Countries	30	30	30	25	25	25

Note: The dependent variable is *Fiscal Consolidation* in year *t*. The method of estimation is two-stage least squares. Huber robust standard errors (shown in parentheses) are clustered at the country level. Columns (1)–(3) show estimates for the whole sample. Columns (4)–(6) show estimates for the sub-sample that excludes France, Germany, Japan, the United Kingdom, and the United States. * Significantly different from zero at the 10 percent level, ** 5 percent level, and *** 1 percent level.

Consider the estimates shown in column (1) of Table 6. In that column, one can see that (i) the coefficient on year t GDP growth is -0.08 and has a standard error of 0.02 , and (ii) the coefficient on the year $t - 1$ to t temperature change is 0.01 and has a standard error of 0.04 . While one can reject the null hypothesis that the coefficient on GDP growth is equal to zero at the 1 percent level (p -value 0.00), one cannot reject the null hypothesis that the coefficient on temperature changes is equal to zero at the conventional significance levels (p -value 0.85). Note that in column (1) of Table 6, GDP growth is instrumented with the GDP growth rate of trading partners and the international commodity price index, both variables from Vegh and Vuletin (2015). For comparison purposes, recall that in the reduced form, the coefficient on temperature changes is -0.07 and has a standard error of 0.04 (see column (1) of Table 1). In terms of the size of these coefficients, the estimated reduced-form effect is about seven times larger than the direct effect. In terms of standard errors, these are about the same for the estimated reduced-form and the estimated direct effect. Not rejecting the null of zero direct effect of temperature changes on fiscal consolidations is thus due to a small coefficient, and not due to a large standard error.

In column (2) of Table 6, one can see that (i) the coefficient on GDP growth is -0.09 and has a standard error of 0.05 , and (ii) the coefficient on GDP growth of trading partners is 0.04 and has a standard error of 0.17 . While one can reject the null hypothesis that the coefficient on GDP growth is equal to zero at the 10 percent level (p -value 0.07), one cannot reject the null hypothesis that the coefficient on GDP growth of trading partners is equal to zero at the conventional significance levels (p -value 0.81). In column (2) of Table 6, the instruments for GDP growth are temperature changes and the international commodity price index. For comparison purposes, recall that in reduced form, the coefficient on GDP growth of trading partners is -0.25 and has a standard error of 0.08 (see column (1) of Table 1). In terms of the size of the coefficients, in absolute value, the reduced-form effect is about six times larger than the direct effect. In terms of standard errors, the standard error on the estimated direct effect is about twice as large as the reduced-form effect. Not rejecting the null hypothesis of a zero direct effect of GDP growth of trading partners on fiscal consolidations is thus mostly due to a smaller estimated coefficient, and not so much to a larger standard error.

Column (3) of Table 6 shows that (i) the coefficient on GDP growth is -0.07 and has a standard error of 0.03 , and (ii) the coefficient on the international commodity price index is -0.005 and has a standard error of 0.012 . While one can reject the null hypothesis that the coefficient on GDP growth is equal to zero at the 5 percent level (p -value 0.02), one cannot reject the null hypothesis that the coefficient on the international commodity price index is zero at the conventional significance levels (p -value 0.69). In column (3) of Table 6, the instruments for GDP growth are temperature changes and the GDP growth rate of trading partners. For comparison, in the reduced form, the estimated coefficient on GDP growth of trading partners is -0.017 and has a standard error of 0.007 (see column (1) of Table 3). In terms of the size of the estimated coefficients, in absolute value, the reduced-form effect is about three times larger than the direct effect. In terms of standard errors, the standard error on the estimated direct effect is about twice as large as the reduced-form effect. Not rejecting the null hypothesis of a zero direct effect of international commodity prices on fiscal consolidations is thus mostly due to a smaller estimated coefficient, and not so much to a larger standard error.

6.1.4. A Dynamic Simultaneous System of Equations

The dynamic version of the simultaneous system of two equations is:

- (1) $FiscalConsolidation_{it} = a_i + b_t + \alpha GDPGrowth_{it} + \Gamma_1 X_{it-1} + \Gamma_2 X_{it-2} + u_{it}$
- (2) $GDPGrowth_{it} = c_i + d_t + \beta FiscalConsolidation_{it} + \Pi_1 X_{it-1} + \Pi_2 X_{it-2} + \Theta Z_{it} + e_{it}$

where subscript i refers to country i and t refers to year t . In the above system, X_{it-1} and X_{it-2} are vectors that include GDP growth and fiscal consolidations in years $t - 1$ and $t - 2$, respectively. Z are the same instrumental variables for GDP growth as in Section 6.1.1.

The main purpose of writing out a dynamic version of the simultaneous system of two equations is to facilitate comparison to the literature that has estimated dynamic effects of fiscal consolidations on GDP growth. I have provided estimates using local projection methods. Local projection methods are widely used in the empirical macro-literature. A recent paper by Plagborg-Møller and Wolf (2021) showed that local projections and VARs estimate the same impulse responses. Montiel and Plagborg-Møller (2021) showed that lag-augmented local projections yield standard errors that are asymptotically valid.

In this section, I only discuss estimates of the contemporaneous effect that GDP growth in year t has on fiscal consolidation in year t . From now on, I refer to these estimates as the contemporaneous effect of GDP growth on fiscal consolidations over a horizon of one year. In the next section, I will discuss estimates over horizons that exceed one year, e.g., two and three years.

The dynamics in the model above follow an autoregressive process of order 2. I have specified the dynamics of the model as such in order to facilitate comparison to previous literature, i.e., Guajardo et al. (2014) and Carrière-Swallow et al. (2021), where Equation (2) is specified as an AR(2) model. To examine the sensitivity of the results to lag selection, I also report estimates from a more parsimonious AR(1) model.

Including X_{it-1} and X_{it-2} as controls means that the estimated coefficients β and α are the effects of “shocks”. Throughout their papers, Guajardo et al. (2014) and Carrière-Swallow et al. (2021) used the term “fiscal consolidation shocks”. The term fiscal consolidation shock simply means that the innovation in the fiscal consolidation variable in year t is not forecastable by past events. As such, a fiscal consolidation shock is not necessarily exogenous to contemporaneous GDP growth, although that was the assumption made in previous literature that estimated the effects of fiscal consolidations on GDP growth using the IMF’s data on action-based fiscal consolidations. That is, literature that used the IMF’s fiscal consolidation variable estimated Equation (2) by OLS, assuming that, in Equation (1), $\alpha = 0$. If in the dynamic model specified above $\alpha \neq 0$, then OLS of Equation (2) yields an inconsistent estimate of β . If $\alpha < 0$, then $\beta^{OLS} < \beta$.

Table 7 reports instrumental variables’ estimates of α . These estimates are obtained by two-stage least squares estimations of the dynamic panel models. Column (1) of Table 7 reports estimates for Equation (1), exactly as specified in this section. The estimates reported in column (1) are for the largest sample in the available data. One can see that the estimated coefficient on year t GDP growth is -0.08 , with a standard error of 0.03. Thus, one can reject the null hypotheses that the coefficient is equal to zero at the 1 percent significance level (p -value 0.004).

Table 7. Contemporaneous effects of GDP growth on action-based fiscal consolidations (dynamic model).

	Fiscal Consolidation					
	(1)	(2)	(3)	(4)	(5)	(6)
		Whole Sample		Excluding Large Economies		
GDP Growth, t	−0.08 *** (0.03)	−0.07 *** (0.02)	−0.07 *** (0.02)	−0.09 *** (0.03)	−0.08 *** (0.02)	−0.08 *** (0.02)
CONTROL VARIABLES						
Fiscal Consolidations, $t - 1$	0.40 *** (0.05)	0.40 *** (0.05)	0.40 *** (0.05)	0.40 *** (0.05)	0.41 *** (0.05)	0.41 *** (0.05)
Fiscal Consolidations, $t - 2$	−0.00 (0.07)	0.01 (0.06)		−0.00 (0.07)	0.01 (0.07)	
GDP Growth, $t - 1$	0.02 (0.01)			0.02 (0.01)		
GDP Growth, $t - 2$	−0.02 ** (0.01)			−0.02 ** (0.01)		
Hansen J, p -value	0.97	0.99	0.98	0.89	0.72	0.83
Cragg–Donald F-Stat	13	20.4	22	9.4	14.7	15.5
Kleibergen–Paap F-Stat	14	13.7	13.8	12.6	12.4	12.1

Table 7. Cont.

	Fiscal Consolidation					
	First-Stage Estimates for GDP Growth, t					
Temperature Change	0.60 *** (0.2)	0.60 *** (0.21)	0.58 *** (0.21)	0.63 *** (0.24)	0.60 *** (0.26)	0.59 ** (0.26)
GDP Growth of Trading Partners	1.64 *** (0.37)	2.16 *** (0.49)	2.04 *** (0.49)	1.61 *** (0.38)	2.14 *** (0.51)	1.99 *** (0.51)
Commodity Price Index	0.14 ** (0.07)	0.20 *** (0.07)	0.22 *** (0.06)	0.13 * (0.07)	0.18 ** (0.07)	0.20 *** (0.06)
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	797	797	824	627	627	649
Countries	29	29	29	24	24	24

Note: The dependent variable is *Fiscal Consolidation* in year t . The method of estimation is two-stage least squares. Huber robust standard errors (shown in parentheses) are clustered at the country level. Columns (1)–(3) show estimates for the whole sample. Columns (4)–(6) show estimates for the sub-sample that excludes France, Germany, Japan, the United Kingdom, and the United States. * Significantly different from zero at the 10 percent level, ** 5 percent level, and *** 1 percent level.

The IV estimate on year t GDP growth that is obtained from the dynamic model is similar to the static model (see Table 4 for comparison). There are fewer observations in Table 7 than in Table 4 because including lags of fiscal consolidations means that the first and second initial observations are lost. Re-estimating the static model on the same number of observations as in column (1) of Table 7 yields a coefficient on year t GDP growth of -0.07 and a standard error of 0.02 .

Concerning the estimated coefficients on the control variables, column (1) of Table 7 shows that the coefficient on year $t - 1$ fiscal consolidations is positive and significantly different from zero. The coefficient on year $t - 2$ fiscal consolidations is quantitatively small and not significantly different from zero at the conventional significance levels. The coefficient on year $t - 1$ GDP growth is positive and not significantly different from zero, while the coefficient on year $t - 2$ GDP growth is negative and significantly different from zero at the 5 percent level.

Column (2) of Table 7 shows estimates of the dynamic model, but without including GDP growth in years $t - 1$ and $t - 2$ as controls. Comparing column (2) to column (1), one can see that including lags of GDP growth has negligible effects on the estimated effect of year t GDP growth on fiscal consolidations. In column (2), the coefficient on year t GDP growth is -0.07 and has a standard error of 0.02 .

Column (3) of Table 7 reports estimates of a more parsimonious model that only includes the first lag of fiscal consolidations as a control variable. The dynamic model in column (3) is an AR(1) model. The motivation for reporting estimates of a more parsimonious model in column (3) is that from columns (1) and (2), one can see that only the coefficient on $t - 1$ fiscal consolidations is significantly different from zero—the coefficient on $t - 2$ fiscal consolidations is not significantly different from zero. In column (3), the coefficient on $t - 1$ fiscal consolidations is 0.40 and has a standard error of 0.05 . Column (3) also shows that the estimated coefficient on GDP growth in year t is negative and significantly different from zero at the conventional significance levels. In column (3), the coefficient on GDP growth in year t is -0.07 and has a standard error of 0.02 .

Comparing columns (1)–(3), one can see that the coefficient on GDP growth in year t is quantitatively similar in size across the different model specifications. This is also the case when re-estimating these models for the sub-sample that excludes the five largest economies (see columns (4)–(6)).

The instrumental variables' estimates in Table 7 are of high quality. Each instrument individually has a significant effect on GDP growth in year t . The Kleibergen–Paap F-statistic is above 10 in all specifications. According to the Hansen J test, one cannot

reject, at the conventional significance levels, the null hypothesis that the instruments are uncorrelated with the second-stage error term.

6.1.5. Integral Multipliers

The fiscal policy literature has computed so-called “integral multipliers” (see, e.g., Ramey (2016)). The dynamic simultaneous equations model for integral multipliers is:

- (1) $FiscalConsolidation(h)_{it} = a_i + b_t + \alpha^h GDPgrowth(h)_{it} + \Gamma_1 X_{it-1} + \Gamma_2 X_{it-2} + u_{it}$
- (2) $GDPgrowth(h)_{it} = c_i + d_t + \beta^h FiscalConsolidation(h)_{it} + \Pi_1 X_{it-1} + \Pi_2 X_{it-2} + \Theta Z(h)_{it} + e_{it}$

where $GDPgrowth(h)_{it}$ is defined as the change in the log of GDP between years $t + h$ and $t - 1$, i.e., $\log GDP_{it+h} - \log GDP_{it-1}$, and $FiscalConsolidation(h)_{it}$ is the sum of fiscal consolidations between years $t + h$ and t . X_{it-1} and X_{it-2} are vectors that include GDP growth and fiscal consolidations in years $t - 1$ and $t - 2$, respectively. $Z(h)_{it}$ is a vector that includes the sum of each instrument between years $t + h$ and $t - 1$. The coefficient α^h is the cumulative effect of GDP growth over h years on the sum of fiscal consolidations over h years. The cumulative effect of the sum of fiscal consolidations over h years on GDP growth over h years is β^h .

For each horizon h , I identified Equation (1) by using $Z(h)$ as instruments for $GDPgrowth(h)$. The instruments are temperature changes, GDP growth of trading partners, and the international commodity price index, computed at the relevant horizon. In the instrumental variables’ regressions of the effect that $GDPgrowth(h)$ has on $FiscalConsolidation(h)$, I computed $Z(h)$ for horizon h as the sum of each instrument between t and $t + h$. I reported instrumental variables’ estimates for $h = 0, 1, 2$ (note that the time unit is years). This is the same number of horizons as in Carrière-Swallow et al. (2021).

Table 8 shows that α^h is negative for all horizons: $h = 0, 1, 2$. Column (1) of Table 8 shows the effect of GDP growth on fiscal consolidations for $h = 0$ (note that this is just a replication of column (1) of Table 7). Column (2) shows the effect of GDP growth on fiscal consolidations for $h = 1$, and column (3) shows the effect for $h = 2$. Thus, one can reject the null hypothesis that the coefficient on $GDPgrowth(h)$ is equal to zero at the 1 percent significance level for $h = 0, 1$, and 2.

Quantitatively, the effects of GDP growth over horizon h on the sum of fiscal consolidations over horizon h are similar in size for $h = 0, 1$, and 2. In columns (1)–(3) of Table 8, the coefficients on $GDPgrowth(h)$ are -0.08 , -0.07 , and -0.07 for $h = 0, 1$, and 2, respectively. The interpretation of these estimated coefficients is as follows:

- i. For $h = 0$, a one standard deviation (equivalent to 4.2) increase in contemporaneous GDP growth over a horizon of one year reduces the magnitude of a fiscal consolidation over a horizon of one year by about 0.3 percent of GDP. This is equivalent to about 0.5 standard deviations.
- ii. For $h = 1$, a one standard deviation (equivalent to 7.2) increase in contemporaneous GDP growth over a horizon of two years reduces the magnitude of a fiscal consolidation over a horizon of two years by 0.5 percent of GDP. This is equivalent to about 0.4 standard deviations.
- iii. For $h = 2$, a one standard deviation (equivalent to 9.5) increase in contemporaneous GDP growth over a horizon of three years reduces the magnitude of a fiscal consolidation over a horizon of three years by 0.7 percent of GDP. This is equivalent to about 0.5 standard deviations.

Results are similar for the sub-sample that excludes the five largest economies (see columns (4)–(6) of Table 8). For the sub-sample that excludes the five largest economies, the estimated coefficients on $GDPgrowth(h)$ are -0.09 , -0.06 , and -0.06 for $h = 0, 1$, and 2, respectively. For each of these estimated coefficients, one can reject the null hypothesis that the coefficient is equal to zero at the 1 percent significance level.

The instrumental variables’ estimates in Table 8 are of high quality. Each instrument individually has a significant effect on $GDPgrowth(h)$ for $h = 0, 1$, and 2. The Kleibergen–Paap F-statistic is above 10 in all specifications. According to the Hansen J test, one cannot

reject, at the conventional significance levels, the hypothesis that the instruments are uncorrelated with the second-stage error term.

Table 8. Contemporaneous effects of GDP growth on action-based fiscal consolidations (at different time horizons).

	Fiscal Consolidation (<i>h</i>)					
	(1) <i>h</i> = 0	(2) <i>h</i> = 1	(3) <i>h</i> = 2	(4) <i>h</i> = 0	(5) <i>h</i> = 1	(6) <i>h</i> = 2
	Whole Sample			Excluding Large Economies		
GDP Growth (<i>h</i>)	−0.08 *** (0.03)	−0.07 *** (0.02)	−0.07 *** (0.02)	−0.09 *** (0.03)	−0.06 *** (0.02)	−0.06 *** (0.02)
CONTROL VARIABLES						
Fiscal Consolidation, <i>t</i> − 1	0.40 *** (0.05)	0.53 ** (0.11)	0.54 *** (0.13)	0.41 *** (0.05)	0.55 ** (0.12)	0.56 *** (0.14)
Fiscal Consolidation, <i>t</i> − 2	−0.00 (0.06)	−0.02 (0.07)	−0.00 (0.07)	−0.00 (0.06)	−0.03 (0.07)	−0.01 (0.07)
GDP Growth, <i>t</i> − 1	0.02 (0.02)	0.01 (0.02)	−0.01 (0.02)	0.02 (0.02)	−0.00 (0.02)	−0.02 (0.02)
GDP Growth, <i>t</i> − 2	−0.02 ** (0.00)	−0.03 ** (0.01)	−0.04 *** (0.02)	−0.02 ** 0	−0.04 *** (0.02)	−0.05 *** (0.02)
	First-Stage Estimates for GDP Growth (<i>h</i>)					
Temperature Change (<i>h</i>)	0.60 *** (0.2)	0.76 ** (0.37)	1.41 *** (0.54)	0.63 *** (0.24)	0.68 * (0.4)	1.21 ** (0.06)
GDP Growth of Trading Partners (<i>h</i>)	1.64 *** (0.37)	2.35 *** (0.56)	3.01 *** (0.73)	1.61 *** (0.38)	2.34 *** (0.57)	3.08 *** (0.74)
Commodity Price Index (<i>h</i>)	0.14 ** (0.07)	0.21 *** (0.07)	0.22 *** (0.08)	0.13 * (0.07)	0.19 *** (0.07)	0.20 *** (0.07)
Cragg–Donald F-Stat	14	16.5	19.1	9.4	15.9	23.01
Kleibergen–Paap F-Stat	13	22.7	32.1	12.6	16.5	23.76
Hansen J, <i>p</i> -value	0.97	0.58	0.42	0.89	0.43	0.27
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	797	768	739	627	603	579
Countries	29	29	29	24	24	24

Note: The dependent variable is *Fiscal Consolidation(h)*, where *h* refers to the horizon, *h* = 0, 1, and 2. The method of estimation is two-stage least squares. Huber robust standard errors (shown in parentheses) are clustered at the country level. Columns (1)–(3) show estimates for the whole sample. Columns (4)–(6) show estimates for the sub-sample that excludes France, Germany, Japan, the United Kingdom, and the United States. * Significantly different from zero at the 10 percent level, ** 5 percent level, and *** 1 percent level.

6.2. The Response of GDP Growth to Fiscal Consolidations

The literature, see in particular Guajardo et al. (2014) and Carrière-Swallow et al. (2021), reported estimates of β^h from OLS estimations of Equation (2). The identifying assumption that previous literature made is that $\alpha^h = 0$ for all *h*. A necessary condition for OLS to yield consistent estimates of β^h is that in the simultaneous system of equations, $\alpha^h = 0$. The instrumental variables' estimates in Table 8 showed that α^h is negative for all horizons, *h* = 0, 1, and 2. Hence, OLS estimation of Equation (2) yields responses of GDP growth (cumulated over horizon *h*) to fiscal consolidation shocks (cumulated over horizon *h*) that are downward-biased for all *h*.⁶

To correct for the downward bias, one needs an instrument when estimating Equation (2). In this section, I report instrumental variables of Equation (2), where the instrument for fiscal consolidations is the residual variation in fiscal consolidation that is not due to GDP growth, i.e., $u^{res} = Fiscalconsolidation(h) - \alpha^{h,IV} GDPGrowth(h)$, where $\alpha^{h,IV}$ is the instrumental variables' estimate of the coefficient on *GDPGrowth(h)* from Table 8. As shown in Section 4, this instrumental variables strategy is immune to the downward bias that arises in OLS estimation of Equation (2) due to $\alpha^h < 0$.

Panel A of Table 9 shows instrumental variables' estimates of the response of GDP growth to fiscal consolidation shocks. Regarding horizons and sample size, Table 9 is

organized exactly as Table 8. That is, column (1) of Table 9 shows the estimated response of GDP growth to contemporaneous fiscal consolidation shocks at horizon $h = 0$. Columns (2) and (3) show the estimates for $h = 1$ and $h = 2$, respectively. Columns (4)–(6) report results for the sub-sample that excludes the five largest economies.

Table 9. Effects of action-based fiscal consolidations on GDP growth.

	GDP Growth (h)					
	(1) $h = 0$	(2) $h = 1$	(3) $h = 2$	(4) $t = 0$	(5) $t = 1$	(6) $t = 2$
	Whole Sample			Excluding Large Economies		
	Panel A: Two-Stage Least Squares					
Fiscal Consolidation (h)	1.84 *** (0.51)	1.46 *** (0.55)	1.38 ** (0.61)	2.00 *** (0.58)	1.61 *** (0.62)	1.52 ** (0.7)
	[Wild Restricted Efficient Cluster Bootstrapped 95% Confidence Interval]					
	[0.72, 2.83]	[0.35, 2.69]	[0.20, 3.08]	[0.67, 3.26]	[0.31, 3.01]	[0.11, 3.49]
	First Stage for Fiscal Consolidation (h)					
u^{res}	0.88 *** (0.03)	0.91 *** (0.03)	0.91 *** (0.03)	0.86 *** (0.03)	0.90 *** (0.04)	0.89 *** (0.05)
Cragg–Donald F-Stat	4253.7	5376.2	4590.9	2878.5	3467.7	2830.2
Kleibergen–Paap F-Stat	807.4	887.7	633	550.9	569.5	378.2
	Panel B: Least Squares					
Fiscal Consolidation (h)	−0.44 *** (0.15)	−0.57 ** (0.22)	−0.74 *** (0.27)	−0.45 ** (0.19)	−0.60 ** (0.24)	−0.80 *** (0.28)
	Controls and Observations in Panels A and B					
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	827	797	767	657	632	607
Countries	30	30	30	25	25	25

Note: The method of estimation in Panel A is two-stage least squares, and in Panel B, least squares. Robust standard errors (shown in parentheses) are clustered at the country level. The 95% confidence intervals [shown in square brackets] are obtained using the wild restricted efficient bootstrap with 1000 draws, clustered at the country level. The auxiliary random variable was drawn from a Rademacher distribution, and bootstrapping was performed over the t -statistic. The endogenous variable in Panel A is $FiscalConsolidation(h)$, where the h refers to the horizon, $h = 0, 1$, and 2 . The instrument is the residual variation in $FiscalConsolidation(h)$ that is not due to $GDPGrowth(h)$, i.e., $u^{\text{res}} = FiscalConsolidation(h) - \alpha^{IV} GDPGrowth(h)$, where α^{IV} is the estimated coefficient on $GDPGrowth(h)$ from Table 8. Additional controls, with estimates not reported, are GDP growth in $t - 1$ and $t - 2$, fiscal consolidations in $t - 1$ and $t - 2$, and computed for each horizon, $h = 0, 1$, and 2 , temperature changes (h), GDP growth of trading partners (h), and the international commodity price index (h). * Significantly different from zero at the 10 percent level, ** 5 percent level, and *** 1 percent level.

According to the instrumental variables’ estimates, fiscal consolidation shocks have a positive effect on GDP growth. Column (1) of Panel A shows that the instrumental variables’ estimate on $FiscalConsolidation(h = 0)$ is 1.84 and has a standard error of 0.51. Thus, one can reject the null hypothesis that this estimated coefficient is equal to zero at the 1 percent significance level (p -value 0.000). Columns (2) and (3) show that the instrumental variables’ estimates for horizons $h = 1$ and $h = 2$ are 1.46 and 1.38, respectively. Thus, one can reject the hypothesis that these effects are equal to zero at the 1 and 5 percent levels, respectively (the p -values are 0.007 and 0.023, respectively). The quantitative interpretation is as follows: over a horizon of one year ($h = 0$), a fiscal consolidation equal to 1 percent of GDP increases GDP by about 1.8 percent, while over a horizon of two years ($h = 1$) and three years ($h = 2$), the cumulative GDP gain relative to the size of the cumulative fiscal consolidation is equal to about 1.5 and 1.4 percent, respectively.

Effects are similar for the sub-sample that excludes the five largest economies (see columns (4)–(6) of Table 9). For the sub-sample that excludes the five largest economies, the estimated coefficients on $FiscalConsolidation(h)$ are 2.00, 1.61, and 1.52, respectively, for $h = 0, 1$, and 2 . For each of these estimated coefficients, one can reject the null hypothesis

that the effect is equal to zero at the conventional significance levels (the p -values are 0.001, 0.010, and 0.030, respectively).

Panel B of Table 9 reports OLS estimates. For all horizons, $h = 0, 1$, and 2 , the OLS estimates of β^h are negative. This means that OLS regressions suggest that fiscal consolidations reduce GDP growth. Specifically, according to the OLS estimates in columns (1)–(3) of Table 9, a fiscal consolidation equal to 1 percent of GDP decreases GDP by about 0.44 percent over a horizon of 1 year ($h = 0$), while over a horizon of two years ($h = 1$) and three years ($h = 2$), GDP decreases by about 0.57 and 0.74 percent, respectively. This is the same result as in previous literature that used OLS to estimate Equation (2). However, as argued above, OLS estimates of the response of GDP growth to action-based fiscal consolidations are downward-biased.

Figure 1 plots the OLS and 2SLS estimates of β^h and their 95 confidence bands. The figure visualizes the results of columns (1)–(3) of Table 9. Recall that the coefficient β^h is the integral multiplier at horizon h , i.e., the ratio of the cumulative change in GDP over the cumulative change in fiscal consolidations. From Figure 1, one can see that the 2SLS estimates of the integral multipliers are positive, while the OLS estimates are negative. OLS and 2SLS estimates of β^h are significantly different. For $h = 0, 1$, and 2 , the 95 percent confidence bands around the OLS and 2SLS estimates of β^h are non-overlapping. The Hausman test rejects the null hypothesis that OLS and 2SLS estimates of β^h are the same, for $h = 0, 1$, and 2 , at the 1 percent level.

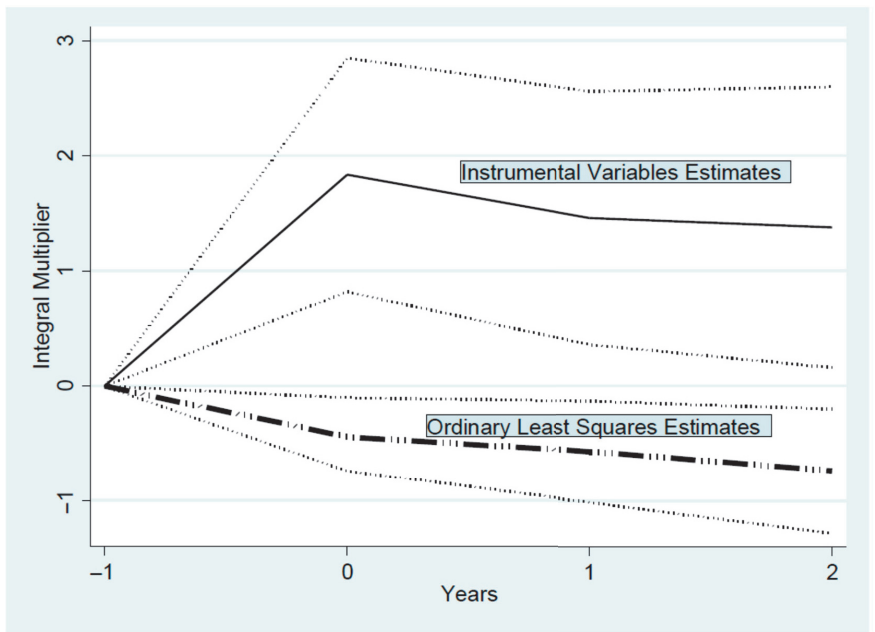


Figure 1. Estimated effect of a 1% of GDP fiscal consolidation on real GDP. Note: The figure shows estimates of the coefficient β^h in Equation (2). The letter h in the superscript refers to the horizon, in years. The solid lines in the above figure are the β^h coefficients obtained from instrumental variables’ regressions. The thick, long-dash-dotted lines are the β^h coefficients obtained from ordinary least squares regressions. The thin, tight-dotted lines are the 95% confidence bands.

6.3. Mechanisms

Table 10 shows that action-based fiscal consolidations have a significant positive effect on total factor productivity. Table 10 also shows that action-based fiscal consolidations have a significant positive effect on investment and the trade balance. The estimated effect of action-based fiscal consolidations on private consumption is positive but not significantly

different from zero at the conventional significance levels. Quantitatively, the estimates in Table 10 show that a fiscal consolidation in year t equal to 1 percent of GDP increases total factor productivity growth in year t by around 0.6 percentage points, increases investment growth in year t by around 2.2 percentage points, and increases the GDP share of net-exports by around 0.8 percentage points. These effects are significantly different from zero at the 5 percent level or higher.

Table 10. Effects of action-based fiscal consolidations on total factor productivity, investment, consumption, and net-exports.

	Total Factor Productivity Growth	Investment Growth	Consumption Growth	Change in Net-Exports GDP Share
	(1)	(2)	(3)	(4)
Panel A: Whole Sample				
Fiscal Consolidation, t	0.62 ** (0.26)	2.34 ** (1.08)	0.48 (0.41)	0.81 *** (0.24)
Cragg–Donald F-Stat	3799.3	3826.8	3999.3	3840.4
Kleibergen–Paap F-Stat	690	689.7	701.9	690.8
Observations	797	797	797	797
Countries	29	29	29	29
Country Fixed Effects	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes
Panel B: Excluding 5 Largest Economies				
Fiscal Consolidation, t	0.62 ** (0.28)	2.35 ** (1.15)	0.42 (0.42)	0.82 *** (0.25)
Cragg–Donald F-Stat	2963.2	2989.8	3146.3	2997.2
Kleibergen–Paap F-Stat	603.3	609.5	622	608.5
Observations	627	627	627	627
Countries	24	29	29	29
Country Fixed Effects	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes

Note: The dependent variable in column (1) is the total factor productivity growth, column (2) investment growth, column (3) private consumption growth, and column (4) the year $t - 1$ to t change in the GDP share of net-exports. The method of estimation is two-stage least squares. Panel A shows estimates for the whole sample. Panel B shows estimates for the sub-sample that excludes France, Germany, Japan, the United Kingdom, and the United States. Huber robust standard errors (shown in parentheses) are clustered at the country level. Additional controls, with estimates not reported, are the dependent variable in $t - 1$ and $t - 2$, GDP growth in $t - 1$ and $t - 2$, and fiscal consolidations in $t - 1$ and $t - 2$. * Significantly different from zero at the 10 percent level, ** 5 percent level, and *** 1 percent level.

To shed more light on the mechanisms, in particular regarding the role of prices, in Table 11, I report the effects that fiscal consolidations have on inflation and the real exchange rate. If prices are very sticky, then one should see an insignificant effect of fiscal consolidations on inflation. If, on the other hand, prices are flexible, then one should see a decrease in inflation. Table 11 shows that fiscal consolidations have a significant negative effect on inflation. The real exchange rate depreciates significantly. Specifically, from Table 11, one can see that a fiscal consolidation in year t equal to 1 percent of GDP decreases the growth rate of the GDP deflator in year t by around 2.3 percentage points, and decreases the year t growth rate of the real exchange rate by around 1.0 percentage points. These effects are significantly different from zero at the 5 percent level or higher. Overall, the results do not support a (New-)Keynesian argument that there may be negative effects of fiscal consolidations on GDP growth.

Table 11. Effects of action-based fiscal consolidations on inflation and the real exchange rate.

	Inflation	Inflation	Real Exchange Rate	Real Exchange Rate
	(1)	(2)	(3)	(4)
	Whole Sample	Excluding 5 Largest Economies	Whole Sample	Excluding 5 Largest Economies
Fiscal Consolidation, <i>t</i>	−2.33 *** (0.75)	−2.44 *** (0.82)	−0.95 ** (0.46)	−1.10 ** (0.48)
Cragg–Donald F-Stat	3793.9	2976.4	3604.4	2758.9
Kleibergen–Paap F-Stat	676.1	600.4	918	893.9
Observations	797	627	674	517
Countries	29	25	25	20
Country Fixed Effects	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes

Note: The dependent variable in columns (1) and (2) is the growth rate of the GDP price deflator, and in columns (3) and (4), the growth rate of the real exchange rate. The method of estimation is two-stage least squares. Columns (1) and (3) show estimates for the whole sample. Columns (2) and (4) show estimates for the sub-sample that excludes France, Germany, Japan, the United Kingdom, and the United States. Huber robust standard errors (shown in parentheses) are clustered at the country level. Additional controls, with estimates not reported, are the dependent variable in *t* − 1 and *t* − 2, GDP growth in *t* − 1 and *t* − 2, and fiscal consolidations in *t* − 1 and *t* − 2. * Significantly different from zero at the 10 percent level, ** 5 percent level, and *** 1 percent level.

7. Conclusions

The latest IMF Fiscal Monitor states that: “durable and credible fiscal consolidation is needed to reestablish sound public finances, to build budgetary space for priority investments, and to deal with future shocks” (International Monetary Fund 2024, p. xi). This call for fiscal consolidation is understandable. In advanced economies, on average, general government debt in 2023 was around 110 percent of the GDP. For the world, on average, general government debt was around 93 percent. These are very high levels of debt. If governments choose to reduce budget deficits, what should one expect to be the effect on GDP growth?

The empirical results presented in this paper suggested that GDP growth would increase if governments were to implement discretionary fiscal policies that reduce budget deficits. Instrumental variables’ estimates showed that action-based fiscal consolidations have a significant positive effect on GDP growth. Action-based fiscal consolidations significantly increase investment, boost the trade balance, and lead to an increase in total factor productivity. A political economy explanation for why the effects of action-based fiscal consolidations on economic growth are positive is that only those politicians who have less (no) present-bias will choose to implement them.

Previous literature, based on OLS, found negative effects of action-based fiscal consolidations on GDP growth. I confirmed this result for the most up-to-date data on GDP growth: OLS regressions yielded negative effects of action-based fiscal consolidations on GDP growth. However, OLS estimates of the effects of action-based fiscal consolidations on GDP growth were downward-biased. This is because GDP growth has a negative contemporaneous effect on the magnitude of fiscal consolidations. My instrumental variables’ estimates of the effects of action-based fiscal consolidations on GDP growth do not suffer from this negative reverse causality bias.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jrfm17050194/s1>.

Funding: This research received no external funding.

Data Availability Statement: Data are available from the sources cited.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

Table A1. Descriptive statistics.

Variable	Source	Mean	Std. Dev.	Obs.
Fiscal Consolidation	DeVries et al. (2011); David and Leigh (2018)	0.32	0.74	1016
Tax-based	DeVries et al. (2011); David and Leigh (2018)	0.15	0.44	984
Expenditure-based	DeVries et al. (2011); David and Leigh (2018)	0.16	0.44	984
Temperature Change	FAOSTAT (2021)	0.68	0.6	994
Temperature Change	Dell et al. (2012)	0.03	0.65	745
GDP Growth of Trading Partners	Vegh and Vuletin (2015)	0.66	0.51	913
Commodity Price Index	Vegh and Vuletin (2015)	0.83	3.62	908
Commodity Price Index	Gruss and Kebhaj (2019)	100.5	3.99	995
Terms of Trade	World World Bank (2021)	99.6	18.5	1012
GDP Growth	PWT version 10.0	3.41	4.24	1016
Total Factor Productivity Growth	PWT version 10.0	0.39	2.11	1016
Investment Growth	PWT version 10.0	3.46	10.7	1016
Consumption Growth	PWT version 10.0	3.31	3.75	1016
Change in GDP Share of Net-Exports	PWT version 10.0	−0.02	3.27	1016
Inflation	PWT version 10.0	2.82	10.15	1016
Real Effective Exchange Rate Growth	World Bank (2021)	−0.08	617	836
GDP Share of Tax Revenues	Vegh and Vuletin (2015)	21.8	8.8	793
GDP Share of Government Expenditures	PWT version 10.0	16	4.3	1016

- Notes**
- ¹ For temperature, examples of papers that document a significant effect on GDP growth are Dell et al. (2012), Burke et al. (2015b), and Gallic and Vermandel (2020). In IV estimations of the contemporaneous effects of annual GDP growth on tax rates, Vegh and Vuletin (2015) showed that GDP growth of trading partners and an international commodity price index are relevant instruments for GDP growth. There are many other papers that have documented a significant effect of these variables on GDP, such as Kose (2002), Acemoglu et al. (2008), Brueckner et al. (2012), and Fernandez et al. (2017).
- ² The narrative approach was pioneered by Friedman and Schwartz (1963) and has since been applied to monetary policy, fiscal policy, and macroprudential policy. See Ramey (2016) for a discussion of recent papers that have applied a narrative approach to various types of economic policy.
- ³ See <https://ideas.repec.org/r/imf/imfwpa/2011-128.html> (accessed on 15 January 2024).
- ⁴ For more details on how these variables are constructed and justification for why they are exogenous to GDP growth, see page 351 of Vegh and Vuletin (2015).
- ⁵ Vegh and Vuletin (2015) also used the US real interest rate as an instrument for GDP growth. I do not report results for the US real interest rate because my panel models include time fixed effects. The US real interest rate is perfectly collinear with the time fixed effects. I have estimated panel models without time fixed effects and included the US real interest rate on the right-hand side. I found a significant positive effect of the US real interest rate on the magnitude of action-based fiscal consolidations in the sample that excludes the US economy: a one percentage point increase in the US real interest rate in year t increases the magnitude of an action-based fiscal consolidation by about 0.03 percent of GDP.
- ⁶ Carrière-Swallow et al. (2021) included, in addition to the first and second lags of GDP growth and fiscal consolidations, current and lagged values of a commodity price index. Carrière-Swallow et al. (2021) did not include temperature changes or GDP growth of trading partners in Equation (2). Note that even with a commodity price index as a control variable in Equation (2), OLS of Equation (2) still yields a downward-biased estimate of β^h if, in Equation (1), $\alpha^h < 0$.

References

Acemoglu, Daron, Simon Johnson, James Robinson, and Pierre Yared. 2008. Income and Democracy. *American Economic Review* 98: 808–42. [CrossRef]

Alesina, Alberto, Carlo Favero, and Francesco Giavazzi. 2015. The output effect of fiscal consolidation plans. *Journal of International Economics* 96: S19–S42. [CrossRef]

Alesina, Alberto, Carlo Favero, and Francesco Giavazzi. 2018. What Do We Know about the Effects of Austerity? *AEA Papers and Proceedings* 108: 524–30. [CrossRef]

Alesina, Alberto, Carlo Favero, and Francesco Giavazzi. 2019a. Effects of Austerity: Expenditure- and Tax-based Approaches. *Journal of Economic Perspectives* 33: 141–62. [CrossRef]

- Alesina, Alberto, Carlo Favero, and Francesco Giavazzi. 2019b. *Austerity: When It Works and When It Doesn't*. Princeton: Princeton University Press.
- Alesina, Alberto, Roberto Perotti, and Jose Tavares. 1998. The Political Economy of Fiscal Adjustments. *Brookings Papers on Economic Activity* 1998: 197–266. [CrossRef]
- Barnichon, Regis, and Davide Debortoli. 2022. Understanding the Size of the Government Spending Multiplier: It's in the Sign. *Review of Economic Studies* 89: 87–117. [CrossRef]
- Blanchard, Oliver, and Roberto Perotti. 2002. An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *Quarterly Journal of Economics* 117: 1329–68. [CrossRef]
- Blanchard, Oliver. 1990. Comment: Can Severe Fiscal Contractions Be Expansionary? Tales of Two Small European Countries. *NBER Macroeconomics Annual* 5: 75–122. [CrossRef]
- Brueckner, Markus. 2012. An instrumental variables approach to estimating tax revenue elasticities: Evidence from Sub-Saharan Africa. *Journal of Development Economics* 98: 220–27. [CrossRef]
- Brueckner, Markus, Alberto Chong, and Mark Gradstein. 2012. Estimating the permanent income elasticity of government expenditures: Evidence on Wagner's law based on oil price shocks. *Journal of Public Economics* 96: 1025–35. [CrossRef]
- Burke, Marshall, Solomon Hsiang, and Edward Miguel. 2015b. Global non-linear effect of temperature on economic production. *Nature* 527: 235–39. [CrossRef]
- Cloyne, James, Oscar Jorda, and Alan Taylor. 2020. *Decomposing the Fiscal Multiplier*. NBER Working Paper 26939. Cambridge: National Bureau of Economic Research.
- Carrière-Swallow, Yan, Antonio C. David, and Daniel Leigh. 2021. Macroeconomic Effects of Fiscal Consolidation in Emerging Economies: New Narrative Evidence from Latin America and the Caribbean. *Journal of Money, Credit and Banking* 6: 1313–35. [CrossRef]
- David, Antonio, and Daniel Leigh. 2018. *A New Action-Based Dataset of Fiscal Consolidation in Latin America and the Caribbean*. IMF Working Papers 18/94. Washington, DC: International Monetary Fund.
- Dell, Melissa, Ben Jones, and Ben Olken. 2012. Temperature Shocks and Economic Growth: Evidence from the Last Half Century. *American Economic Journal: Macroeconomics* 4: 66–95. [CrossRef]
- DeVries, Paul, Jaime Guajardo, Daniel Leigh, and Andrea Pescatori. 2011. *A New Action-Based Dataset of Fiscal Consolidation*. IMF Working Paper No. 11/128. Washington, DC: International Monetary Fund.
- Food and Agriculture Organization Corporate Statistical Database. 2021. Available online: <http://www.fao.org/faostat/en/#data/QC> (accessed on 2 April 2024).
- Fernandez, Andres, Stephanie Schmitt-Grohe, and Martin Uribe. 2017. World shocks, world prices, and business cycles: An Empirical Investigation. *Journal of International Economics* 108: S2–S14. [CrossRef]
- Friedman, Milton, and Anna Jacobson Schwartz. 1963. *A Monetary History of the United States, 1867–1960*. Princeton: Princeton University Press.
- Gallie, Ewen, and Gallier Vermandel. 2020. Weather Shocks. *European Economic Review* 124: 103409. [CrossRef]
- Giavazzi, Francesco, and Marco Pagano. 1990. Can Severe Fiscal Contractions Be Expansionary? Tales of Two Small European Countries. *NBER Macroeconomics Annual* 5: 75–122. [CrossRef]
- Gruss, Bertrand, and Suhaib Kebhaj. 2019. *Commodity Terms of Trade: A New Database*. IMF Working Paper No. 19/21. Washington, DC: International Monetary Fund.
- Guajardo, Jaime, Daniel Leigh, and Andrea Pescatori. 2014. Expansionary Austerity? International Evidence. *Journal of the European Economic Association* 12: 949–68. [CrossRef]
- Hausman, Jerry, Whitney Newey, and William Taylor. 1987. Efficient Estimation and Identification of Simultaneous Equation Models with Covariance Restrictions. *Econometrica* 55: 849–74. [CrossRef]
- International Monetary Fund. 2024. *Fiscal Monitor*. Washington, DC: International Monetary Fund.
- Jordà, Oscar, and Alan Taylor. 2016. The Time for Austerity: Estimating the Average Treatment Effect of Fiscal Policy. *Economic Journal* 126: 219–55. [CrossRef]
- Kose, Ayhan. 2002. Explaining business cycles in small open economies: How much do world prices matter? *Journal of International Economics* 56: 299–327. [CrossRef]
- Montiel, Jose, and Mikkel Plagborg-Møller. 2021. Local Projection Inference is Simpler and More Robust Than You Think. *Econometrica* 89: 1789–823. [CrossRef]
- Mountford, Andrew, and Harald Uhlig. 2009. What are the Effects of Fiscal Policy Shocks? *Journal of Applied Econometrics* 24: 960–92. [CrossRef]
- Plagborg-Møller, Mikkel, and Christian Wolf. 2021. Local Projections and VARs Estimate the Same Impulse Responses. *Econometrica* 89: 955–80. [CrossRef]
- Pappa, Evi, Rana Sajedi, and Eugenia Vella. 2015. Fiscal consolidation with tax evasion and corruption. *Journal of International Economics* 1: 56–75. [CrossRef]
- Ramey, Valerie. 2011. Identifying Government Spending Shocks: It's All in the Timing. *The Quarterly Journal of Economics* 126: 51–102. [CrossRef]
- Ramey, Valerie. 2016. Macroeconomic Shocks and Their Propagation. In *Handbook of Macroeconomics*. Edited by John B. Taylor and Harald Uhlig. Amsterdam: North Holland, vol. 2A, chp. 2. pp. 71–162.

- Ramey, Valerie. 2019. Ten Years After the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research? *Journal of Economic Perspectives* 33: 89–114. [CrossRef]
- Ramey, Valerie, and Sarah Zubairy. 2018. Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data. *Journal of Political Economy* 126: 850–901. [CrossRef]
- Romer, Christina, and David Romer. 2010. The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks. *American Economic Review* 100: 763–801. [CrossRef]
- Stock, James, and Mark Watson. 2018. Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments. *Economic Journal* 128: 917–48. [CrossRef]
- Vegh, Carlos, and Guillermo Vuletin. 2015. How Is Tax Policy Conducted over the Business Cycle? *American Economic Journal: Economic Policy* 7: 327–70. [CrossRef]
- World Bank. 2021. World Development Indicators. Online Database. Available online: <https://databank.worldbank.org/source/world-development-indicators> (accessed on 19 February 2021).
- Yared, Piere. 2019. Rising Government Debt: Causes and Solutions for a Decades-Old Trend. *Journal of Economic Perspectives* 33: 115–40. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

Testing and Ranking of Asset Pricing Models Using the GRS Statistic

Mark J. Kamstra ^{1,*},[†] and Ruoyao Shi ^{2,†}

¹ Schulich School of Business, Room N204-C, York University, 4700 Keele St., Toronto, ON M3J 1P3, Canada

² Department of Economics, University of California Riverside, 900 University Avenue, Riverside, CA 92521, USA; ruoyao.shi@ucr.edu

* Correspondence: mkamstra@yorku.ca; Tel.: +1-(416)-736-2100 (ext. 33302)

[†] These authors contributed equally to this work.

Abstract: We clear up an ambiguity in the statement of the GRS statistic by providing the correct formula of the GRS statistic and the first proof of its F-distribution in the general multiple-factor case. Casual generalization of the Sharpe-ratio-based interpretation of the single-factor GRS statistic to the multiple-portfolio case makes experts in asset pricing studies susceptible to an incorrect formula. We illustrate the consequences of using the incorrect formulas that the ambiguity in GRS leads to—over-rejecting and misranking asset pricing models. In addition, we suggest a new approach to ranking models using the GRS statistic p -value.

Keywords: GRS; asset pricing; CAPM; multivariate test; portfolio efficiency; Sharpe ratio; over-rejection; model ranking

JEL Classification: G12; G14

1. Introduction

In an influential paper, Gibbons et al. (1989) developed and analyzed a test of the ex ante mean-variance efficiency of portfolios. This test statistic is now widely used to evaluate asset pricing models and has also been exploited to rank competing models. For the single factor case, Gibbons et al. (1989) carefully developed the statistic in a linear regression model (hereafter referred to as the GRS statistic or test), derived its small-sample F distribution, investigated its power properties, and highlighted its significance in asset pricing theory by purveying an alternative interpretation involving the Sharpe ratio (Sharpe 1966)—the excess return to a portfolio per unit of risk (or volatility, measured by standard deviation)—which is a key measure of portfolio efficiency. For the multiple factor case, however, Gibbons et al. (1989, sec. 7), were ambiguous on how the statistic should be constructed.

The solution to the portfolio optimization problem that yields the Sharpe ratio has us estimate a variance–covariance matrix of the portfolio excess returns, but the equivalence of the GRS statistic and the F test statistic relies on this matrix arising in the projection of the test asset returns on the column space of the asset pricing factors, not as a variance–covariance matrix. Unfortunately, Gibbons et al. (1989) used equivocal language to describe this matrix, referring to it as a “variance-covariance matrix”, and this has apparently caused confusion about the function of the GRS statistic, which is further exacerbated by the fact that the small-sample F distribution wrongfully conjures up a degrees-of-freedom (d.f. hereafter) adjustment that is improper in this case. This has led to the application of a very common incorrect formula that, paradoxically, is more likely to be used by financial economists, the experts in the field, than by someone who focuses only on the statistical aspects of the problem.¹ We find that using the incorrect formula, which we will refer to for conciseness as $\hat{W}$ below, leads to (i) a test statistic that does not follow the F distribution

Citation: Kamstra, Mark J., and Ruoyao Shi. 2024. Testing and Ranking of Asset Pricing Models Using the GRS Statistic. *Journal of Risk and Financial Management* 17: 168. <https://doi.org/10.3390/jrfm17040168>

Academic Editor: Thanasis Stengos

Received: 21 March 2024

Revised: 4 April 2024

Accepted: 8 April 2024

Published: 19 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

as prescribed and over-rejects the null hypothesis of portfolio efficiency; and (ii) smaller models often being favored over larger ones when the statistic is used to rank asset pricing models. This error comes from mixing terms that fall out of portfolio optimization with a statistical object that comes from the small-sample F test derivation.

The main contribution of our paper is to clear up the ambiguity in the calculation of the GRS statistic and highlight (both theoretically and empirically) issues that arise from the use of $\hat{W}$ and two related and popularly used statistics, one which folds in a second degree of freedom error (we will refer to this as $\tilde{W}$), and the asymptotic χ^2 version often used to replace the GRS test.² The asymptotic χ^2 and $\tilde{W}$ implementations result in much higher model rejection rates than the correct GRS statistic, most notably for the asymptotic χ^2 test, even with 50 years of monthly data. The use of an incorrect implementation of the GRS statistic also results in inconsistent model rankings across $\hat{W}$, $\tilde{W}$ and the correct calculation of the GRS statistic, with 40 or even 50 years of data. Finally, we propose a new methodology for the ranking of competing asset pricing models, making use of test p -values rather than the raw GRS statistic values, meant to properly internalize the model sizes. While a determination of statistically significantly different model performance is valuable, often researchers are simply attempting to rank models. Our approach is a computationally straightforward approach to answering this question.

We will adopt the notation in Gibbons et al. (1989, sec. 7), whenever possible. The proofs of the theoretical results and the details of the empirical results are in the Appendix A.

2. The GRS Test for Multiple Factors

The proofs of all the claims in this section can be found in Section A. The problem is to test the mean-variance efficiency of L portfolios utilizing another type of N assets (known as test assets).

We start with a linear regression model:

$$\tilde{r}_{it} = \delta_{i0} + \delta_i' \tilde{r}_{pt} + \tilde{\eta}_{it}, \quad \forall i = 1, \dots, N, \text{ and } t = 1, \dots, T, \quad (1)$$

where $\tilde{r}_{it}$ denotes the excess return on test asset i in period t , the L -vector of portfolio excess returns $\tilde{r}_{pt}$ serves as factors, and $\tilde{\eta}_{it}$ denotes the disturbance. Mean-variance efficiency of the L portfolios implies (Sharpe 1964)

$$H_0 : \delta_{i0} = 0, \quad \forall i = 1, \dots, N. \quad (2)$$

Lemma 1 (Joint F test). Let $\tilde{r}_p \equiv [\tilde{r}_{p1}, \dots, \tilde{r}_{pT}]'$, $\tilde{r}_p \equiv T^{-1} \sum_{t=1}^T \tilde{r}_{pt}$, and let $\hat{\delta}_0$ be the ordinary least squares (OLS) estimator of $\delta_0 \equiv (\delta_{10}, \dots, \delta_{N0})'$; also let $\hat{\eta}_t \equiv (\hat{\eta}_{1t}, \dots, \hat{\eta}_{Nt})'$ be the OLS residuals of model (1). We follow Gibbons et al. (1989) to assume that the disturbance $\tilde{\eta}_t \equiv (\tilde{\eta}_{1t}, \dots, \tilde{\eta}_{Nt})'$ is independent from the factors $\tilde{r}_{pt}$ and has a joint normal distribution³ with mean zero and nonsingular variance–covariance matrix Σ and is iid over t . Define

$$\tilde{\Omega}_* \equiv \frac{1}{T} \sum_{t=1}^T \tilde{r}_{pt} \tilde{r}_{pt}', \quad (3)$$

$$\hat{\Sigma} \equiv \frac{1}{T-L-1} \sum_{t=1}^T \hat{\eta}_t \hat{\eta}_t'. \quad (4)$$

Then, the F statistic

$$\tilde{W}_* \equiv \frac{T(T-N-L)}{N(T-L-1)} \left(1 - \tilde{r}_p' \tilde{\Omega}_*^{-1} \tilde{r}_p \right) \hat{\delta}_0' \hat{\Sigma}^{-1} \hat{\delta}_0 \quad (5)$$

follows the $F_{N,T-N-L}$ distribution under H_0 .

From a purely statistical perspective, Lemma 1 is all we need for testing the implication (2) of mean-variance efficiency, which is just the usual joint F test of zero intercepts in a linear regression system.⁴

The economic interpretation of the GRS test, however, is better understood via another implication of mean-variance efficiency— θ_{N+L}^* , the Sharpe ratio of the optimal portfolio consisting of the L portfolios and the N test assets, equals θ_p^* , the Sharpe ratio of the L portfolios alone (Gibbons et al. 1989).

We consider a general portfolio optimization problem that yields the Sharpe ratio. Let $\tilde{r}$ denote a vector of excess returns of K assets ($K \geq 1$), and let $\mu_{\tilde{r}}$ and $\Omega_{\tilde{r}}$ be their ex ante mean vector and variance–covariance matrix, respectively. Let m be the target mean excess return and ω be a vector of K asset weights. The optimal portfolio weights ω^* solve

$$\min_{\omega} \omega' \Omega_{\tilde{r}} \omega, \quad \text{subject to} \quad \omega' \mu_{\tilde{r}} = m.$$

The square of the Sharpe ratio of the optimal portfolio composed of these K assets, therefore, is

$$\theta^{*2} \equiv \left(\frac{m}{\sqrt{\omega^{*'} \Omega_{\tilde{r}} \omega^*}} \right)^2 = \mu_{\tilde{r}}' \Omega_{\tilde{r}}^{-1} \mu_{\tilde{r}},$$

in which the variance–covariance matrix $\Omega_{\tilde{r}}$ of the K assets plays a central role. Applying this general result twice, we obtain that

$$W \equiv \left(\frac{\sqrt{1 + \theta_{N+L}^{*2}}}{\sqrt{1 + \theta_p^{*2}}} \right)^2 - 1 = \left(1 + \mu_{\tilde{r}_p}' \Omega^{-1} \mu_{\tilde{r}_p} \right)^{-1} \delta_0' \Sigma^{-1} \delta_0, \quad (6)$$

where Ω is the variance–covariance matrix of L portfolio excess returns $\tilde{r}_{pt}$ and Σ is that of the disturbances $\tilde{\eta}_t$. So, $W = 0$ if the L portfolios are efficient, and this is the basis of the GRS test in asset pricing theory.

Theorem 1 (Generalized GRS statistic). *Define*

$$\tilde{\Omega} \equiv \frac{1}{T} \sum_{t=1}^T (\tilde{r}_{pt} - \bar{r}_p) (\tilde{r}_{pt} - \bar{r}_p)' = \frac{1}{T} \sum_{t=1}^T \tilde{r}_{pt} \tilde{r}_{pt}' - \bar{r}_p \bar{r}_p' \quad (7)$$

and the generalized GRS statistic

$$\tilde{W} \equiv \frac{T(T - N - L)}{N(T - L - 1)} \left(1 + \bar{r}_p' \tilde{\Omega}^{-1} \bar{r}_p \right)^{-1} \hat{\delta}_0' \hat{\Sigma}^{-1} \hat{\delta}_0. \quad (8)$$

Then, $\tilde{W} = \tilde{W}_*$, and therefore under the conditions of Lemma 1, $\tilde{W}$ follows the $F_{N, T-N-L}$ distribution under the H_0 .

Theorem 1 connects the statistical perspective to the economic interpretation of the GRS test, because $\tilde{W}$ equals the F statistic $\tilde{W}_*$ and can be regarded as a sample analog of W —replace Ω in Equation (6) with its maximum likelihood estimator (MLE) $\tilde{\Omega}$, Σ with its unbiased estimator $\hat{\Sigma}$, $\mu_{\tilde{r}_p}$ with $\bar{r}_p$, δ_0 with $\hat{\delta}_0$, and pre-multiply the ratio $\frac{T(T-N-L)}{N(T-L-1)}$, then one obtains $\tilde{W}$ in Equation (8).

Common Mistakes and Consequences

$\tilde{W}$ equals the original GRS statistic when $L = 1$. For the $L > 1$ case, however, Gibbons et al. (1989, p. 1146) gave a statistic $\hat{W}$, almost identical to $\tilde{W}$, but instead of $\tilde{\Omega}$, they

prescribed “sample variance-covariance matrix” $\hat{\Omega}$ without giving its explicit formula. Since the sample variance-covariance matrix customarily entails a d.f. adjustment, i.e.,

$$\hat{\Omega} \equiv \frac{1}{T-1} \sum_{t=1}^T (\tilde{r}_{pt} - \bar{r}_p)(\tilde{r}_{pt} - \bar{r}_p)' = \frac{T}{T-1} \tilde{\Omega}, \quad (9)$$

this would cause $\hat{W}$ to differ from $\tilde{W}$, and therefore Theorem 1 implies that $\hat{W}$ does not follow the $F_{N,T-N-L}$ distribution as prescribed.

This incorrect GRS statistic $\hat{W}$ inflicts two consequences on empirical asset pricing studies. First, it over-rejects mean-variance efficiency when gauged against the $F_{N,T-N-L}$ distribution, because the ratio between $\hat{W}$ and $\tilde{W}$ is always larger than 1. Second, it misranks competing asset pricing models, because the ratio between $\hat{W}$ and $\tilde{W}$ tends to be disproportionately larger for models with more factors.

The significance of the GRS statistic in recent financial studies, as Fama and French (2015) advocate, resides in the ranking of competing asset pricing models, rather than testing them. Using even the correct GRS statistics to rank models, albeit its portfolio optimization interpretation, is subject to a familiar critique akin to the use of R^2 for linear regression model comparison. Instead, the p -values associated with $\tilde{W}$ in respective $F_{N,T-N-L}$ distributions are a statistically sound metric for this purpose, as they internalize the difference in the second d.f.

The implementations of the GRS test found in popular user-defined software packages, such as `GRS.test` in R and `grstest` and `grstest2` in Stata, not only use $\hat{\Omega}$ when computing the GRS statistic, but also fold in additional errors. Results for the R formula are labeled with $\hat{W}$ in the rest of this paper.

The asymptotic χ^2 test is frequently recommended as an alternative of the F test. One might think that when T is large, the d.f. issue we point out here can be circumvented by using the asymptotic χ^2 test. Unfortunately, we find that the commonly used χ^2 test statistics also over-reject for any sample size, especially if the number of test assets N or the number of factors L is large. In addition, they erroneously favor smaller models to an extent worse than $\hat{W}$.

3. Empirical Results

We use portfolios of test assets borrowed from Fama and French (2015, 2016) to show that over-rejection and misranking of $\hat{W}$, $\check{W}$ and χ^2 relative to $\tilde{W}$ is empirically significant, even remarkable in many cases. To summarize our empirical findings (which are detailed in Appendix B and below), the $\check{W}$ misapplication of the GRS statistic and the asymptotic χ^2 both result in much higher model rejection rates than the correct GRS statistic, most notably for the asymptotic χ^2 test, even with 50 years of monthly data, and also result in scrambled model rankings, with 40 or even 50 years of data, most notably for the $\check{W}$ version of the GRS statistic. While the F test is asymptotically equivalent to χ^2 , typical sample sizes available in financial markets research are not large enough to make this approximation innocuous. The exact F test construction is also the most conservative test, resulting in less over-rejection of the null hypothesis when the null is correct, even with highly non-normal return data, by measure of the bootstrap resampling experiments we perform.

In Table 1, we highlight the over-rejection issue, with five-year windows. This span of data shows serious over-rejection of asset pricing models from the application of the alternative formulations of the GRS test. Results for the largest number of test assets we considered, 32, are displayed in the first three rows of the table, the results for cases with 25 test assets follow in rows four through thirteen, the 17 test assets of the industry portfolios follow in row fourteen, and the remaining six rows present results for sets of 10 test assets. In Table 1, we use a total of 53 five-year overlapping windows starting from 1963 and consider six different asset pricing models, meaning that we have 318 cases for which an asset pricing model might be rejected, for each of the 19 sets of test assets.

The asymptotic χ^2 statistic fares the worst relative to the correct GRS statistic among the alternatives; it over-rejects dramatically relative to the correct GRS statistic, 50% of the time on average, faring worse when we consider more test assets. $\tilde{W}$ and $\hat{W}$ over-reject relative to the correct GRS statistic about 10% and 1% of the time on average. $\hat{W}$ is unstable across test assets, however, with a few cases displaying close to 4% over-rejection, and some with no over-rejection.

Table 1. Number and proportion of cases with more rejections relative to the GRS test for five-year windows sampled during 1963–2019, across 19 sets of test assets.

Test Assets Nb of Subsamples = 53	Asymptotic χ^2			$\tilde{W}$			$\hat{W}$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
2 × 4 × 4 MExMEBExINV	289	275	247	4	17	21	0	1	0
2 × 4 × 4 MExMEBExOP	262	246	220	4	14	18	0	0	0
2 × 4 × 4 MExOPxINV	282	238	194	10	27	28	0	3	1
5 × 5 AccrualsxME	222	218	207	7	17	31	0	0	1
5 × 5 BExME	212	191	166	8	21	20	1	0	1
5 × 5 BetaxME	229	221	204	11	16	25	0	1	0
5 × 5 MExOP	238	227	188	9	27	42	0	0	0
5 × 5 MomentumxME	210	166	124	14	29	27	0	0	0
5 × 5 NetIssuexME	216	176	147	9	29	27	0	2	2
5 × 5 RVariancecxME	147	94	65	18	28	12	0	2	0
5 × 5 VariancecxME	137	81	45	26	27	12	0	3	1
5 × 5 BExInv	248	259	245	4	18	16	0	2	1
5 × 5 MExInv	214	189	171	14	17	29	0	0	0
Industry	126	153	152	12	9	22	0	1	0
Book-to-Market Deciles	48	70	67	5	22	21	0	1	1
Investment Deciles	22	30	56	4	4	6	0	0	0
Momentum Deciles	56	60	66	18	14	20	1	1	1
Size Deciles	49	90	68	6	23	27	0	0	2
Operating Profitability Deciles	48	70	65	4	24	15	0	1	0
Average	171.3	160.7	142	9.8	20.2	22.1	0.11	0.95	0.58
Proportion (%)	53.9	50.5	44.6	3.1	6.3	6.9	0.0	0.3	0.2

Notes: (1) The figures are the number of all decisions at the stated significance level for which the test statistic rejects the model when the correct GRS statistic $\tilde{W}$ does not reject, out of a total of 318 possible cases, with the exception of the last row in which the proportion is given. The sample periods of five years are sampled from July 1963 to December 2019, and the number of models tested is 6 for each window, which we sum over to obtain the total number of over-rejections. These windows overlap, adjusted in a rolling window so that all but 1 year of data overlaps with the next sample window. This means that there are 53 samples for the 5-year window. (2) For a detailed description of the factor and test asset construction see Fama and French (2015, 2016).

In Table 2, we consider the model misranking among six models for each of the 53 five-year windows of Table 1. Misranking of models shows similar problems for the χ^2 statistic as we saw for test rejections, with over 40% of the cases displaying misranking of at least one asset pricing model. The $\tilde{W}$ statistic displays much worse performance than test rejections, with close to 60% of the cases displaying misranking, and we see worse performance even for the $\hat{W}$, with close to 3% of the cases misranked.

While the typical Fama and French paper uses 40 or 50 years of data, it is also true that much empirical work uses far less data. Gibbons et al. (1989) noted that issues of stationarity can reasonably constrain the length of a time series used, so that “it is not uncommon to see published work where T is around 60”, Affleck-Graves and McDonald (1989) limited their analysis and simulations to 60 month periods, Ferson and Foerster (1994) studied 60, 120, and 720 monthly observations in their simulation, Rouwenhorst (1999) used five years in subsample analysis, and among recent works that exploited as

little as four or five years of data are Belimam et al. (2018) and Qin (2019). Leite et al. (2018) used as few as 98 months of data, Lewellen et al. (2010) used 168 observations of quarterly data, Choi et al. (2020) performed subsample stability tests using eight years of monthly data, and many studies of emerging economy markets have used 10 to 15 years of monthly data. See, for instance, Alhomaiddi et al. (2019), Alshammari and Goto (2022), Merdad et al. (2015), and Sha and Gao (2019).

One takeaway from these papers is that many situations involving specialized data (like Sha and Gao 2019) and their exploration of mutual fund returns in China) or subsample robustness checks (like Baek and Bilson 2015) are necessarily constrained to shorter samples than fifty or even twenty years, so that the bias from an incorrectly calculated GRS statistic becomes large.

Table 2. Number and proportion of cases with different ranking outcomes from the GRS statistic for five-year windows sampled during 1963–2019, across 19 sets of test assets

Test Assets Nb of Subsamples = 53	Any Model Mis-Ranked		Top Model Mis-Ranked	
	$\tilde{W}$	$\widehat{W}$	$\tilde{W}$	$\widehat{W}$
$2 \times 4 \times 4$ MExMEBExINV	26	2	5	0
$2 \times 4 \times 4$ MExMEBExOP	37	1	6	0
$2 \times 4 \times 4$ MExOPxINV	31	1	12	0
5×5 AccrualsxME	36	1	12	0
5×5 BExME	26	1	7	0
5×5 BetaxME	31	1	10	0
5×5 MExOP	33	0	10	0
5×5 MomentumxME	33	0	14	0
5×5 NetIssuexME	27	4	9	2
5×5 RVariancexME	32	0	14	0
5×5 VariancexME	34	1	10	0
5×5 BExInv	35	3	6	0
5×5 MExInv	30	3	9	0
Industry	29	2	4	0
Book-to-Market Deciles	28	0	4	0
Investment Deciles	24	2	4	0
Momentum Deciles	30	2	5	0
Size Deciles	27	3	5	0
Operating Profitability Deciles	30	0	6	0
Average	30.47	1.42	8.00	0.11
Proportion (%)	57.5	2.7	15.1	0.2

Notes: (1) The figures are the number of all misrankings by the test statistic value across models relative to the correct GRS statistic ranking, out of a total of 53 possible cases, with the exception of the last row in which the proportion is given. The sample periods of five years are sampled over July 1963 to December 2019, and the number of models ranked is 6 for each window. These windows overlap, adjusted in a rolling window, so that all but 1 year of data overlaps with the next sample window. This means that there are 53 samples for the 5-year window. (2) For a detailed description of the factor and test asset construction see Fama and French (2015, 2016).

4. Concluding Remarks

The GRS statistic of Gibbons et al. (1989), developed to provide a test of the ex ante mean-variance efficiency of portfolios and more recently exploited to rank competing models, can be easily implemented incorrectly due to an ambiguity in the presentation of the multivariate form of the test in Gibbons et al. (1989). This presentation suggests a degree-of-freedom-adjusted unbiased variance–covariance matrix estimator $\hat{\Omega}$ of the portfolio excess returns used in the small-sample GRS F test. Indeed, the portfolio optimization problem naturally has us estimate the variance–covariance matrix Ω of the portfolio excess

returns, but the equivalence of the GRS statistic to the F test relies on $\tilde{\Omega}$, which arises in the projection matrix of the test asset returns on the column space of the asset pricing factors, not as a variance–covariance matrix. Paradoxically, this error is clearly visible when turning a blind eye to the economic interpretation of the GRS statistic and taking a purely statistical approach. Although an unbiased estimator $\hat{\Omega}$ appears intuitive in the context of portfolio optimization, it does not yield a correct small-sample exact F test.

Further complicating this ambiguity, Cochrane (2005) presented the GRS statistic omitting a degree-of-freedom adjustment in the calculation of the variance–covariance matrix of the regression residuals, Σ .⁵ Perhaps an outcome of Cochrane (2005), there is an implementation of the GRS statistic in an R package, which we label $\tilde{W}$, that omits the d.f. adjustment when estimating Σ but fails to pre-multiply the correct ratio.⁶ It has also become common in the field to ignore the F distribution completely and employ an asymptotic χ^2 approximation in place of the F test.

The main results for both the asymptotic χ^2 and $\tilde{W}$ implementations is much higher model rejection rates than the correct GRS statistic, most notably for the asymptotic χ^2 test, and they also result in scrambled model rankings. Further, the F distribution is inherently pertinent to small-sample exact tests, where one should make a point of computing the d.f. correctly. For this reason, we recommend the exact F test construction with its attendant F distribution, for both testing and ranking of asset pricing models. The exact F test construction is also the most conservative test, resulting in less over-rejection of the null hypothesis when the null is correct, even with highly non-normal return data.

Another result of this research inquiry is that we provided the first proof of the F-distribution of this test for the general multi-factor case and we recommended a new ranking method, making use of the *p*-value rather than the raw GRS statistic value. Although ranking by the values of the GRS statistic has a desirable economic intuition attached to it, the applied researcher taking advantage of this must recognize that this ranking is statistically as unsound as favoring a regression model with the highest R^2 .

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jrfm17040168/s1>.

Author Contributions: Conceptualization, methodology, review, editing, software, validation, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition were all shared equally. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Social Sciences and Humanities Research Council of Canada, grant number 510991.

Data Availability Statement: Data used is available through Ken French’s data library.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proofs and Details for the Results in Section 2

The following two lemmas are used in the proof of Lemma 1.

Lemma A1. If a random vector Y and a random matrix W satisfy: (i) $Y \sim \mathcal{N}_d(\mu, \Sigma)$, the d dimensional normal distribution; (ii) $W \sim \mathcal{W}_d(f, \Sigma)$, the $d \times d$ dimensional Wishart distribution; and (iii) $Y \perp W$. Then, given Hotelling’s *T*-squared defined as $T^2 \equiv f(Y - \mu)'W^{-1}(Y - \mu)$, we have $F \equiv \frac{f-d+1}{fd}T^2 \sim F_{d,f-d+1}$.

Lemma A2 (Sherman–Morrison formula). Suppose A is an invertible $L \times L$ matrix and u and v are $L \times 1$ vectors. If $A + uv'$ is invertible, then $(A + uv')^{-1} = A^{-1} - \frac{A^{-1}uv'A^{-1}}{1+v'A^{-1}u}$.

Lemma A1 is a standard result in multivariate statistics (see, e.g., Anderson 2003, Theorem 5.2.2), and Lemma A2 is a standard result in linear algebra (see, e.g., Bartlett 1951, p. 107).

Proof of Lemma 1. The proof proceeds in three steps.

Step 1. In this step, we show that under the null hypothesis (2),

$$\sqrt{T(1 - \tilde{r}'_p \tilde{\Omega}_*^{-1} \tilde{r}_p)} \hat{\delta}_0 \sim \mathcal{N}_N(0, \Sigma), \quad (\text{A1})$$

where $\tilde{\Omega}_*$ is defined in Equation (3).

Let ℓ_T denote a $T \times 1$ vector with every element being one, and let I_T denote the $T \times T$ identity matrix. Define $P_{p,T} = \tilde{r}_p (\tilde{r}'_p \tilde{r}_p)^{-1} \tilde{r}'_p$ as the $T \times T$ projection matrix (onto the column space of $\tilde{r}_p$) and its $T \times T$ complement matrix $Q_{p,T} = I_T - P_{p,T}$. It is a standard result (e.g., Hayashi 2000, pp. 18–19) that the OLS estimator of δ_{i0} satisfies $\hat{\delta}_{i0} - \delta_{i0} = (\ell'_T Q_{p,T} \ell_T)^{-1} \ell'_T Q_{p,T} \tilde{\eta}_i$, where $\tilde{\eta}_i \equiv (\tilde{\eta}_{i1}, \dots, \tilde{\eta}_{iT})'$ for $\forall i = 1, \dots, N$. Since $\tilde{\eta}_{it}$ has a normal distribution, and let σ_{ii}^2 denote the (i, i) entry of Σ , then it is a standard result (e.g., Hayashi (2000, Sec. 1.3) that $\sqrt{\ell'_T Q_{p,T} \ell_T} (\hat{\delta}_{i0} - \delta_{i0}) \sim \mathcal{N}_1(0, \sigma_{ii}^2)$. It then only takes some algebra to show that

$$\sqrt{\ell'_T Q_{p,T} \ell_T} (\hat{\delta}_0 - \delta_0) \sim \mathcal{N}_N(0, \Sigma). \quad (\text{A2})$$

Now, let us take a closer look at $\ell'_T Q_{p,T} \ell_T$:

$$\begin{aligned} \ell'_T Q_{p,T} \ell_T &= \ell'_T \ell_T - \ell'_T \tilde{r}_p (\tilde{r}'_p \tilde{r}_p)^{-1} \tilde{r}'_p \ell_T \\ &= T - \left(\sum_{t=1}^T \tilde{r}'_{pt} \right) \left(\sum_{t=1}^T \tilde{r}_{pt} \tilde{r}'_{pt} \right)^{-1} \left(\sum_{t=1}^T \tilde{r}_{pt} \right) \\ &= T - T \left(\frac{1}{T} \sum_{t=1}^T \tilde{r}'_{pt} \right) \left(\frac{1}{T} \sum_{t=1}^T \tilde{r}_{pt} \tilde{r}'_{pt} \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^T \tilde{r}_{pt} \right) \\ &= T(1 - \tilde{r}'_p \tilde{\Omega}_*^{-1} \tilde{r}_p). \end{aligned} \quad (\text{A3})$$

Recall that $\delta_0 = 0$ under the null hypothesis (2), so Equation (A2) and (A3) together imply (A1), the claim of Step 1.

Step 2. In this step, we will show that $\hat{\delta}_0 \perp \hat{\Sigma}$ and

$$(T - L - 1) \hat{\Sigma} \sim \mathcal{W}_N(T - L - 1, \Sigma). \quad (\text{A4})$$

Let $X = [\ell_T, \tilde{r}_p]$ denote the $T \times (L + 1)$ design matrix of Equation (1). Define the $T \times T$ projection matrix $P = X(X'X)^{-1}X'$ and its complement $Q = I_T - P$. Let $\tilde{\eta} = [\tilde{\eta}_1, \dots, \tilde{\eta}_N]$ denote the $T \times N$ matrix of all disturbances in Equation (1). Then, by the standard results of the OLS estimators with normal disturbances (e.g., Hayashi (2000, Sec. 1.3), we have $\hat{\delta}_0 \perp \hat{\Sigma}$ and $(T - L - 1) \hat{\Sigma} = \sum_{t=1}^T \hat{\eta}_t \hat{\eta}'_t = \tilde{\eta}' Q \tilde{\eta} = \tilde{\eta}' U D U' \tilde{\eta}$, where the last equality holds by the singular value decomposition of Q , in which U is a $T \times T$ unitary matrix, and D is a $T \times T$ diagonal matrix with $T - L - 1$ diagonal entries being ones and the rest being zeros. Since we assume that the rows of $\tilde{\eta}$ are mutually independent and follow the $\mathcal{N}_N(0, \Sigma)$ distribution, the rows of $U' \tilde{\eta}$ are also mutually independent and follow the $\mathcal{N}_N(0, \Sigma)$ distribution. This further implies that $\tilde{\eta}' U D U' \tilde{\eta}$ has the same distribution as such sum $S = \sum_{j=1}^{T-L-1} \xi_j \xi'_j$, where ξ_j are mutually independent and $\xi_j \sim \mathcal{N}_N(0, \Sigma)$ ($j = 1, \dots, T - L - 1$). By construction, the distribution of S is the Wishart distribution $\mathcal{W}_N(T - L - 1, \Sigma)$. This proves the claim of Step 2.

Step 3. In this step, we apply Lemma A1 to the results of Steps 1 and 2. After some simple algebra, we obtain $\tilde{W}_* \sim F_{N, T-N-L}$ with $\tilde{W}_*$ defined in Equation (5). This completes the proof of Lemma 1. $\square$

Derivation of Equation (6). Gibbons et al. (1989) derives this, in Section 6, for the $L = 1$ case, and here we provide the derivation for the general $L \geq 1$ case. We start by considering a general portfolio optimization problem that yields the Sharpe ratio—mean excess return to a portfolio per unit of volatility (standard deviation)—of the optimal portfolio consisting of given assets. Let $\tilde{r}$ denote a vector of excess returns of K assets ($K \geq 1$), and let $\mu_{\tilde{r}}$ and $\Omega_{\tilde{r}}$ be their ex ante mean vector and variance–covariance matrix, respectively. Let m be the target mean excess return and ω be a vector of K asset weights. The optimal portfolio weights ω^* solve

$$\min_{\omega} \omega' \Omega_{\tilde{r}} \omega, \quad \text{subject to} \quad \omega' \mu_{\tilde{r}} = m.$$

The first order conditions for this problems are $\omega^* = \varphi \Omega_{\tilde{r}}^{-1} \mu_{\tilde{r}}$ and $\varphi = m / (\mu_{\tilde{r}}' \Omega_{\tilde{r}}^{-1} \mu_{\tilde{r}})$, where φ is the Lagrange multiplier. The squared Sharpe ratio of the optimal portfolio consisting of these K assets is, therefore,

$$\theta^{*2} \equiv \left(\frac{m}{\sqrt{\omega^{*'} \Omega_{\tilde{r}} \omega^*}} \right)^2 = \mu_{\tilde{r}}' \Omega_{\tilde{r}}^{-1} \mu_{\tilde{r}},$$

in which the variance–covariance matrix $\Omega_{\tilde{r}}$ plays a central role.

Applying this general result, we know that when the constituent assets are the L portfolios, the squared Sharpe ratio is

$$\theta_p^{*2} = \mu_{\tilde{r}_p}' \Omega^{-1} \mu_{\tilde{r}_p}. \quad (\text{A5})$$

When the constituent assets include both the N test assets and the L portfolios, the squared Sharpe ratio is $\theta_{N+L}^{*2} = \mu_{\tilde{r}_{N+L}}' \Omega_{\tilde{r}_{N+L}}^{-1} \mu_{\tilde{r}_{N+L}}$, where $\mu_{\tilde{r}_{N+L}} \equiv \left(\mu_{\tilde{r}_N}', \mu_{\tilde{r}_p}' \right)'$,

$$\Omega_{\tilde{r}_{N+L}} \equiv \begin{bmatrix} \delta \Omega \delta' + \Sigma & \delta \Omega \\ \Omega \delta' & \Omega \end{bmatrix}, \quad (\text{A6})$$

and $\delta \equiv [\delta_1, \dots, \delta_N]'$ with δ_i being the slope coefficient in model (1). Equation (A6) holds because we can rewrite the variance–covariance matrix of the N test assets and their covariance matrix with the L portfolios using Ω , Σ and δ (in the same way as $\hat{V}$ on p. 1143 and eq. (24) in Gibbons et al. 1989). Applying the inverse formula for a block matrix and noticing the relationship between $\mu_{\tilde{r}_N}$ and $\mu_{\tilde{r}_p}$ implied by model (1), we obtain

$$\theta_{N+L}^{*2} = \theta_p^{*2} + \delta_0' \Sigma^{-1} \delta_0, \quad (\text{A7})$$

which is essentially the same as Equations (22) and (23) in MacKinlay and Richardson (1991). This, together with Equation (A5) and simple algebra, further implies Equation (6).

Proof of Theorem 1. Based on Lemma 1, we only need to show that $\tilde{W}_*$ defined in Equation (5) equals $\tilde{W}$ in Equation (8). By comparing Equation (3) and (7), we see that $\tilde{\Omega} = \tilde{\Omega}_* - \tilde{r}_p \tilde{r}_p'$, so it suffices to show that

$$1 - \tilde{r}_p' \tilde{\Omega}_*^{-1} \tilde{r}_p = \left(1 + \tilde{r}_p' \tilde{\Omega}^{-1} \tilde{r}_p \right)^{-1} = \left[1 + \tilde{r}_p' \left(\tilde{\Omega}_* - \tilde{r}_p \tilde{r}_p' \right)^{-1} \tilde{r}_p \right]^{-1}. \quad (\text{A8})$$

Applying Lemma A2 with $A = \tilde{\Omega}_*$, $u = \tilde{r}_p$ and $v = -\tilde{r}_p$, we get $\left(\tilde{\Omega}_* - \tilde{r}_p \tilde{r}_p' \right)^{-1} = \tilde{\Omega}_*^{-1} + \frac{\tilde{\Omega}_*^{-1} \tilde{r}_p \tilde{r}_p' \tilde{\Omega}_*^{-1}}{1 - \tilde{r}_p' \tilde{\Omega}_*^{-1} \tilde{r}_p}$, which implies that $1 + \tilde{r}_p' \left(\tilde{\Omega}_* - \tilde{r}_p \tilde{r}_p' \right)^{-1} \tilde{r}_p = 1 + \tilde{r}_p' \tilde{\Omega}_*^{-1} \tilde{r}_p + \frac{(\tilde{r}_p' \tilde{\Omega}_*^{-1} \tilde{r}_p)^2}{1 - \tilde{r}_p' \tilde{\Omega}_*^{-1} \tilde{r}_p} = \left(1 - \tilde{r}_p' \tilde{\Omega}_*^{-1} \tilde{r}_p \right)^{-1}$, which further immediately implies Equation (A8). This completes the proof of Theorem 1. $\square$

Original GRS statistic when $L = 1$. When $L = 1$, $\tilde{\Omega}$ equals to $\frac{1}{T} \sum_{t=1}^T \tilde{r}_{pt}^2 - \bar{r}_p^2 = s_p^2$, the sample variance of $\tilde{r}_{pt}$ **without** d.f. defined by Gibbons et al. (1989, p. 1124). So, $\tilde{W}$ equals the original GRS statistic when $L = 1$.

Over-rejection of $\hat{W}$. Take the ratio between $\hat{W}$ and $\tilde{W}$, then by the relationship between $\hat{\Omega}$ and $\tilde{\Omega}$ in Equation (9), we get

$$\frac{\hat{W}}{\tilde{W}} = \frac{1 + \tilde{r}_p' \tilde{\Omega}^{-1} \tilde{r}_p}{1 + \frac{T-1}{T} \tilde{r}_p' \tilde{\Omega}^{-1} \tilde{r}_p}, \quad (\text{A9})$$

which measures how much the incorrect formula inflates the GRS statistic. Define a function $g(x) = \frac{1+x}{1+\frac{T-1}{T}x}$. Since the first-order derivative of this function is $g'(x) = \frac{1/T}{(1+\frac{T-1}{T}x)^2} > 0$, we know that $g(x)$ is a monotonically increasing function of x . This, combined with the facts that $g(0) = 1$ and $\tilde{r}_p' \tilde{\Omega}^{-1} \tilde{r}_p > 0$, implies that $\hat{W}/\tilde{W} > 1$. As a result, when $\hat{W}$ is gauged against the $F_{N,T-N-L}$, the distribution of $\tilde{W}$, it will over-reject the null hypothesis of mean-variance efficiency of the L portfolios.

Model misranking by $\hat{W}$. Some back-of-the-envelope calculation shows that $\tilde{r}_p' \tilde{\Omega}^{-1} \tilde{r}_p$ tends to be larger for models with more factors. To see this, let $\mu_{\tilde{r}_p}$ denote the mean vector of $\tilde{r}_{pt}$ as above, then by the central limit theorem, we have $\sqrt{T}(\tilde{r}_p - \mu_{\tilde{r}_p}) \xrightarrow{d} \mathcal{N}(0, \Omega)$; and by the law of large numbers, we have $\tilde{\Omega} \xrightarrow{p} \Omega$. These two results imply that $T(\tilde{r}_p - \mu_{\tilde{r}_p})' \tilde{\Omega}^{-1} (\tilde{r}_p - \mu_{\tilde{r}_p}) \xrightarrow{d} \chi_L^2$. Note that $E(\chi_L^2) = L$, so this in turn implies that for fixed T , the mean of $\tilde{r}_p' \tilde{\Omega}^{-1} \tilde{r}_p$ is approximately $E(\tilde{r}_p' \tilde{\Omega}^{-1} \tilde{r}_p) \approx \frac{L}{T} + \mu_{\tilde{r}_p}' \Omega^{-1} \mu_{\tilde{r}_p}$, where $\mu_{\tilde{r}_p}' \Omega^{-1} \mu_{\tilde{r}_p}$ is expected to increase with L since the dimensions of both $\mu_{\tilde{r}_p}$ and Ω increase with L . As a result, the random variable $\tilde{r}_p' \tilde{\Omega}^{-1} \tilde{r}_p$ tends to increase with L on average.⁷

Combined with Equation (A9), this means that the ratio $\hat{W}/\tilde{W}$ tends to be larger for larger models; that is, smaller models tend to be disproportionately favored if the incorrect GRS statistic $\hat{W}$ is used to rank models, compared to the ranking based on the correct GRS statistic $\tilde{W}$.

Additional errors in software packages. The R package `GRS.test` computes two different GRS statistics, see Kim (2022). One (function `GRS.test`) uses the unbiased estimators $\hat{\Omega}$ and $\hat{\Sigma}$ at the same time; the other (function `GRS.MLtest`) uses the MLEs $\tilde{\Omega}$ and $\tilde{\Sigma} \equiv \frac{1}{T} \sum_{t=1}^T \hat{\eta}_t \hat{\eta}_t'$ at the same time. The former is just $\tilde{W}$, and we denote the latter as

$$\tilde{W} \equiv \frac{T(T-N-L)}{N(T-L-1)} \left(1 + \tilde{r}_p' \tilde{\Omega}^{-1} \tilde{r}_p\right)^{-1} \delta_0' \tilde{\Sigma}^{-1} \delta_0, \text{ and note } \tilde{W} = \frac{T}{T-L-1} \tilde{W}. \quad (\text{A10})$$

These two statistics are both incorrect and clearly stem from the interpretation of Ω and Σ as variance-covariance matrices in Gibbons et al. (1989).

Because of the relationship between $\tilde{W}$ and $\tilde{W}$ in Equation (A10), similar analysis as in that for $\hat{W}$ indicates the same over-rejection and misranking problems for $\tilde{W}$ as well, even to a worse extent than $\hat{W}$ for typical data in empirical asset pricing studies.

The Stata packages `grstest` and `grstest2`, composed by different contributors,⁸ make use of $\hat{\Omega}$ and further compound this error by estimating Σ as $\frac{1}{T-1} \sum_{t=1}^T \hat{\eta}_t \hat{\eta}_t'$ and pre-multiplying the ratio $\frac{T-N-L}{N}$ instead of $\frac{T(T-N-L)}{N(T-L-1)}$. The result is a statistic that is difficult to justify and different from all those we discussed above.⁹

Asymptotic χ^2 test. First note that the distribution of the correct GRS statistic $\tilde{W}$, when multiplied by N , converges to the χ_N^2 distribution as $T \rightarrow \infty$.¹⁰ So, comparing $N\tilde{W}$ with the critical value from the χ_N^2 distribution, rather than comparing $\tilde{W}$ with the $F_{N,T-N-L}$ critical value, is by itself an asymptotically valid χ^2 test. The commonly used χ^2 test statistics in

empirical asset pricing research deviate from $N\tilde{W}$, and the deviations are all positive,¹¹ so they will over-reject compared to $N\tilde{W}$.¹² We find that the χ^2 statistics misrank models more often even than $\tilde{W}$ in our empirical studies, but we do not report the model ranking results for χ^2 , because they do not have an intuitive economic interpretation in the model ranking context, and therefore are not commonly used for this purpose. The misrankings of these χ^2 can be easily shown by a similar analysis as for $\tilde{W}$ and $\tilde{W}$ and are therefore skipped here.

Appendix B. Details of the Empirical Results

We now turn to some empirical examples, focusing on how the different implementations of the GRS statistic, as well as the asymptotic χ^2 statistic, compare to the correct calculation, based on model testing and ranking outcomes, borrowing from Fama and French (2015, 2016) the choice of asset pricing models and the choice of test assets. The models we consider include the CAPM, the Fama–French three-factor model, two variations of a four-factor model, the Fama–French five-factor model and a six-factor model that includes momentum. The test assets we explore include 5×5 sortings based on market capitalization and various anomaly variables including operating profitability, return volatility, residual volatility, accruals and so on, up to as many as $32 (2 \times 4 \times 4)$ portfolio sortings. We also explore decile portfolio sortings based on size, operating profitability, momentum, book-to-market and investment. The number of test assets used in empirical work is commonly as large as 25, as we see in Fama and French (2015, 2016), though many studies use 30 to over 50 test assets. See, for instance, Lewellen et al. (2010), Kroencke (2017), Demaj et al. (2018), and Kleiberger and Zhan (2020). Recently, asset pricing models have typically contained at least four or five factors, though six are also commonly seen. See, for instance, Barillas and Shanken (2018), Fama and French (2018), Kan et al. (2024) or Hanauer (2020). Given the state of the literature, our choice of test assets and factors sits comfortably amidst the typical empirical asset pricing applications.

We use data retrieved from the French data library, and we consider five, ten, fifteen, twenty, twenty-five, forty, and fifty-year periods drawn from 1963–2019 for our consideration of up to six factors in the competing asset pricing models, and from 1926–2019 for our consideration of up to four factors in the competing asset pricing models.¹³ We limit our sample window to no less than five years of monthly data because few studies use less than 60 observations; Gibbons et al. (1989) note that issues of stationarity can reasonably constrain the length of a time series used, so that “it is not uncommon to published work where T is around 60”, Affleck-Graves and McDonald (1989) limited their analysis and simulations to 60 month periods; Ferson and Foerster (1994) studied 60, 120, and 720 monthly observations in their simulation study; Rouwenhorst (1999) used five years in sub-sample analysis; and among recent work that exploited as few as four or five years of data are Belimam et al. (2018), and Qin (2019). Leite et al. (2018) used as few as 98 months of data, Lewellen, Nagel, and Shanken (2010) used 168 observations of quarterly data, Choi et al. (2020) performed sub-sample stability tests using eight years of monthly data, and many studies of emerging economy markets have used ten to fifteen years of monthly data. See, for instance, Alhomaidi et al. (2019), Alshammari and Goto (2022), Merdad et al. (2015), and Sha and Gao (2019).

Our primary results, found in Tables A1–A7, make use of the full sample available to us by partitioning the data sample into overlapping periods. For instance, at the five year horizon over 1963–2019, we form five-year windows starting in 1963 and every year following, so that the first window extends from July 1963 to June 1967, the second from January 1964 to December 1968, January 1965 to December 1969, and so on, resulting in 53 five-year overlapping windows. For each of these windows over the period 1963–2019, we use 19 sets of test assets, listed in the first column of Table A1, and six competing asset pricing models. These models are the CAPM, the Fama–French three-factor model, four and five-factor models, as well as a six-factor model including momentum, all as considered in Fama and French (2015, 2016).

Appendix B.1. Results for Five-Year Windows

We first present a small subset of our empirical findings in Tables A1–A3. For convenience, Tables A1 and A2 replicate Tables 1 and 2 from the main text, and here we discuss them in greater depth. In these tables, we consider five-year windows, the minimum span of data the GRS statistic is commonly applied to. This short span of data shows the most serious over-rejection of asset pricing models from the application of the alternative formulations of the GRS test, as well as the highest frequency of misrankings relative to the correct formulation of the GRS statistic. We present evidence for longer spans of data, up to 50 year windows, in Tables A4–A7, and discuss them in Appendix B.2.

In Table A1, we present the number of excess test rejections at the 1%, 5%, and 10% levels, relative to the correct GRS statistic $\tilde{W}$, for each of the alternative test statistics, the asymptotic χ^2 , $\hat{W}$ and $\check{W}$. The results for the largest number of test assets we considered, 32, are displayed in the first three rows of the table; the results for cases with 25 test assets follow on rows four through thirteen; the 17 test assets of the industry portfolios follow on row fourteen; and the remaining five rows present results for sets of 10 test assets. In Table A1, we use a total of 53 five-year overlapping windows starting from 1963 and consider six different asset pricing models, meaning that we have 318 cases for which an asset pricing model might be rejected, for each of the 19 sets of test assets.

Table A1. Number and proportion of subsamples with more rejections relative to the GRS test for five year windows sampled during 1963–2019, across 19 sets of test assets.

Test Assets Nb of Subsamples = 53	Asymptotic χ^2			$\check{W}$			$\hat{W}$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
2 × 4 × 4 MExMEBExINV	289	275	247	4	17	21	0	1	0
2 × 4 × 4 MExMEBExOP	262	246	220	4	14	18	0	0	0
2 × 4 × 4 MExOPxINV	282	238	194	10	27	28	0	3	1
5 × 5 AccrualsxME	222	218	207	7	17	31	0	0	1
5 × 5 BExME	212	191	166	8	21	20	1	0	1
5 × 5 BetaxME	229	221	204	11	16	25	0	1	0
5 × 5 MExOP	238	227	188	9	27	42	0	0	0
5 × 5 MomentumxME	210	166	124	14	29	27	0	0	0
5 × 5 NetIssuexME	216	176	147	9	29	27	0	2	2
5 × 5 RVariancecxME	147	94	65	18	28	12	0	2	0
5 × 5 VariancecxME	137	81	45	26	27	12	0	3	1
5 × 5 BExInv	248	259	245	4	18	16	0	2	1
5 × 5 MExInv	214	189	171	14	17	29	0	0	0
Industry	126	153	152	12	9	22	0	1	0
Book-to-Market Deciles	48	70	67	5	22	21	0	1	1
Investment Deciles	22	30	56	4	4	6	0	0	0
Momentum Deciles	56	60	66	18	14	20	1	1	1
Size Deciles	49	90	68	6	23	27	0	0	2
Operating Profitability Deciles	48	70	65	4	24	15	0	1	0
Average	171.3	160.7	142	9.8	20.2	22.1	0.11	0.95	0.58
Proportion (%)	53.9	50.5	44.6	3.1	6.3	6.9	0.0	0.3	0.2

Notes: (1) The figures are the number of all decisions at the stated significance level for which the test statistic rejects the model when the correct GRS statistic does not reject, out of a total of 318 possible, with the exception of the last row for which the proportion is given. The sample periods of five years are sampled over July 1963 to December 2019 and the number of models tested are 6 for each window, which we sum over to obtain the total number of over-rejections. These windows overlap, adjusted in a rolling window so that all but 1 year of data overlaps with the next sample window. This means that there are 53 samples for the 5 year window. (2) For a detailed description of the factor and test asset construction see Fama and French (2015, 2016).

The asymptotic χ^2 statistic fares the worst relative to the correct GRS statistic among the alternatives, over-rejecting roughly 50% of the time on average relative to $\tilde{W}$, across

the common significance levels of 1%, 5%, and 10%. This over-rejection is worse when we consider more test assets.

The $\tilde{W}$ and $\hat{W}$ do not display patterns related to the number of test assets or significance level, with the $\tilde{W}$ ($\hat{W}$) over-rejecting relative to the correct GRS statistic about 5% (0.2%) of the time on average, across the common significance levels of 1%, 5%, and 10%. The over-rejection of the $\hat{W}$ is unstable across test assets, however, with a few cases displaying close to 6% over-rejection (three times over 53 subsamples), and some with no over-rejection.

In Table A2, we present the number of cases for which each statistic misranks factor models relative to the correct GRS statistic ranking. Here, we rank the six models for each of the 53 five-year windows considered in Table A1. Misranking of models at the five year horizon by the $\tilde{W}$ statistic displays much worse performance than we saw for test rejections, with close to 60% of the cases displaying a misranking, and we saw even worse performance for the $\hat{W}$, with close to 3% of the cases misranked. If we restrict our attention to cases for which the top model is misranked, the $\tilde{W}$ statistic misranks between 10% and 15% of the time, while the $\hat{W}$ misranks the top model less than 0.5% of the time.

Table A2. Number and proportion of subsamples with different ranking outcomes from the GRS statistic for five year windows sampled during 1963–2019, across 19 sets of test assets.

Test Assets Nb of Subsamples = 53	Any Model Mis-Ranked		Top Model Mis-Ranked	
	$\tilde{W}$	$\hat{W}$	$\tilde{W}$	$\hat{W}$
2 × 4 × 4 MExMEBExINV	26	2	5	0
2 × 4 × 4 MExMEBExOP	37	1	6	0
2 × 4 × 4 MExOPxINV	31	1	12	0
5 × 5 AccrualsxME	36	1	12	0
5 × 5 BExME	26	1	7	0
5 × 5 BetaxME	31	1	10	0
5 × 5 MExOP	33	0	10	0
5 × 5 MomentumxME	33	0	14	0
5 × 5 NetIssuexME	27	4	9	2
5 × 5 RVariancexME	32	0	14	0
5 × 5 VariancexME	34	1	10	0
5 × 5 BExInv	35	3	6	0
5 × 5 MExInv	30	3	9	0
Industry	29	2	4	0
Book-to-Market Deciles	28	0	4	0
Investment Deciles	24	2	4	0
Momentum Deciles	30	2	5	0
Size Deciles	27	3	5	0
Operating Profitability Deciles	30	0	6	0
Average	30.47	1.42	8.00	0.11
Proportion (%)	57.5	2.7	15.1	0.2

Notes: (1) The figures are the number of all misrankings from a particular test statistic value across models relative to the GRS statistic ranking, out of a total of 53 possible, with the exception of the last row for which the proportion is given. The sample periods of five years are sampled over July 1963 to December 2019 and the number of models ranked are 6 for each window. These windows overlap, adjusted in a rolling window so that all but 1 year of data overlaps with the next sample window. This means that there are 53 samples for the 5 year window. (2) For detailed description of the factor and test asset construction see Fama and French (2015, 2016).

In Table A3, we present detailed results for the 5 × 5 net share issuance’s crossed with size portfolio test asset set, for five overlapping five-year windows near the end of the 1963–2019 sample period, in order to give the reader a finer sense for the results in Tables A1 and A2. We present the average annualized and raw percentage alpha for each of the six asset pricing models and each window, as well as values of the correct GRS statistic $\tilde{W}$, and the $\tilde{W}$ and $\hat{W}$ statistics. Beside each test statistic value, we report the ranks of the six models, from one to six. A model ranked differently by $\tilde{W}$ or $\hat{W}$ from $\tilde{W}$ is indicated by a † next to the factor label and further identified with the appropriate column’s rank

number being boldfaced. A misranked top model is indicated by an * next to the factor label and further identified with the appropriate row's factor label being bolded. A test method producing different ranks using p -values from test statistic is indicated by a ‡ next to the window period and further identified with a ↓ in the appropriate column. This issue of different ranking using the test statistic versus the p -value will be drawn out below.

Table A3. Summary statistics on factor models and test statistics for five year windows over 2007–2009 financial crisis sample period, for investment 5×5 NetIssuexME test assets.

Date/ Factor Model	Average Annualized /Raw % α	$\tilde{W}$ Statistic /Rank	$\hat{W}$ Statistic /Rank	$\hat{W}$ Statistic /Rank
JAN 2005-DEC 2009				
Mkt	2.63/0.219	0.688/1	0.712/1	0.688/1
Mkt SMB HML	2.39/0.199	0.819/2	0.878/2	0.819/2
Mkt SMB HML UMD	2.37/0.198	0.837/3	0.913/3	0.837/3
Mkt SMB RMW CMA	2.54/0.212	0.910/4	0.992/4	0.912/4
Mkt SMB HML RMW CMA	2.52/0.210	0.997/6	1.108/6	0.999/6
Mkt SMB HML RMW CMA UMD	2.57/0.214	0.972/5	1.100/5	0.974/5
JAN 2006-DEC 2010 ‡				
Mkt †	3.23/0.269	1.114/3	↓	1.114/3
Mkt SMB HML †	2.32/0.193	1.078/2	1.155/2	1.078/1
Mkt SMB HML UMD †	2.41/0.201	1.256/6	1.371/5	1.257/6
Mkt SMB RMW CMA *,†	2.64/0.220	1.077/1	1.174/3	1.080/2
Mkt SMB HML RMW CMA	2.65/0.221	1.130/4	1.256/4	1.134/4
Mkt SMB HML RMW CMA UMD †	2.81/0.234	1.253/5	1.418/6	1.257/5
JAN 2007-DEC 2011 ‡				
Mkt †	3.39/0.283	1.329/2	1.375/1	↓
Mkt SMB HML †	3.11/0.259	1.503/4	1.610/3	1.329/1
Mkt SMB HML UMD	3.26/0.272	1.888/5	2.060/5	1.504/4
Mkt SMB RMW CMA *,†	2.89/0.240	1.327/1	1.448/2	1.890/5
Mkt SMB HML RMW CMA †	2.92/0.244	1.478/3	1.643/4	1.332/2
Mkt SMB HML RMW CMA UMD	3.08/0.257	1.947/6	2.204/6	1.484/3
JAN 2008-DEC 2012				
Mkt	4.34/0.361	1.299/2	1.344/2	1.300/2
Mkt SMB HML	3.64/0.303	1.500/3	1.607/3	1.500/3
Mkt SMB HML UMD †	3.77/0.314	1.906/6	2.079/5	1.907/6
Mkt SMB RMW CMA	2.98/0.248	1.161/1	1.267/1	1.165/1
Mkt SMB HML RMW CMA	3.15/0.263	1.577/4	1.753/4	1.583/4
Mkt SMB HML RMW CMA UMD †	3.39/0.282	1.896/5	2.146/6	1.904/5
JAN 2009-DEC 2013 ‡				
Mkt	2.74/0.228	↓	1.678/2	↓
Mkt SMB HML	3.05/0.254	2.938/5	1.736/2	1.681/2
Mkt SMB HML UMD	3.12/0.260	3.324/6	3.148/5	2.946/5
Mkt SMB RMW CMA	2.46/0.205	1.402/1	3.626/6	3.333/6
Mkt SMB HML RMW CMA	3.05/0.254	2.723/3	1.530/1	1.406/1
Mkt SMB HML RMW CMA UMD	2.72/0.227	2.733/4	3.026/3	2.734/3
			3.093/4	2.745/4

Notes: (1) A test method producing different ranks using p -values versus test statistic is indicated by a ‡ in the assets label. This is further identified with a ↓ in the appropriate column. Ranked test statistics different than $\tilde{W}$ test ranked value is indicated by a † on the factor model label. This is further identified with the appropriate column's rank being bolded. Top ranked test statistics different than top ranked $\tilde{W}$ test value are indicated by a * in the assets label. This is further identified by the appropriate row's model being bolded. (2) For a detailed description of the factor and test asset construction see Fama and French (2015, 2016).

What we see in Table A3 is fairly typical across the full set of empirical findings that Tables A1 and A2 are based on; the $\hat{W}$ displays many misrankings, and misrankings of the top model are rare. Although not tabulated, only the asymptotic χ^2 is typically rejecting factor models in this particular small set of examples, so that the $\tilde{W}$, $\hat{W}$, and correct $\tilde{W}$ test

statistic are all consistent with each other. Common in studies that seek to rank models are rankings by the average absolute alpha, whether or not all models are rejected by the GRS test. See for instance Fama and French (2015). As we can see from Table A3, often the model with the smallest average absolute alpha is not top-ranked.

Appendix B.2. Summary Results for Longer Windows

We now present evidence for a longer span of data in Tables A4–A7, using two sets of date windows up to 50 years. In addition to the period 1963–2019 that we considered in Appendix B.1, now we add the period 1926–1962. The pre-1963 period lacks data for factors and test assets built using operating profitability, accruals etc., so we are left with six test assets constructed from book-to-market, size, industry classification, and momentum, and three different factor models, including the CAPM, the Fama–French three-factor model, and a four-factor model including momentum, all as considered in Fama and French (2015, 2016). For this restricted set of test assets and factors, we estimate test rejections and rankings using the entire 1926–2019 sample and we break out results separately from those constructed using the larger set of test assets and models on the 1963–2019 data alone. It is interesting to do this, as the chance of a misranking declines with fewer asset pricing models being considered.

In Table A4, we present the percentage of excess test rejections relative to the correct GRS statistic at each of the 1%, 5%, and 10% levels for each of the alternative test methods, the asymptotic χ^2 , the $\tilde{W}$, and the $\hat{W}$. This is performed for overlapping windows of 5, 10, 15, 20, 25, 40, and 50 years, using data that span either 1963–2019 or 1926–2019. Panel A displays the percentage of over-rejection rates averaged over six asset pricing models and 19 sets of test assets for the period 1963–2019; Panel B displays the same percentages for the period 1926–2019 using the smaller set of three factor models and six sets of test assets.

Table A4. Percentage of subsamples/models with different decision outcomes from the correct GRS test $\tilde{W}$.

Window (Months)	χ^2			$\tilde{W}$			$\hat{W}$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
Panel A: 1963–2019									
60	53.9	50.5	44.6	3.1	6.3	6.9	0.0	0.3	0.2
120	27.4	24.8	20.0	3.2	3.7	3.9	0.1	0.0	0.1
180	16.4	13.0	11.4	2.1	2.1	2.2	0.1	0.0	0.0
240	9.7	8.6	6.3	1.4	1.7	1.2	0.0	0.0	0.0
300	6.5	5.7	4.5	0.9	1.3	1.1	0.1	0.1	0.0
480	3.0	2.9	1.9	0.3	0.7	0.4	0.0	0.0	0.0
600	2.9	0.8	1.6	0.3	0.2	0.3	0.0	0.0	0.0
Panel B: 1926–2019									
60	34.6	36.1	33.2	2.6	4.3	5.3	0.0	0.1	0.1
120	15.4	15.5	12.6	2.2	3.1	2.5	0.0	0.0	0.0
180	8.8	10.3	9.2	1.0	1.4	1.7	0.1	0.0	0.0
240	8.6	7.7	4.7	1.3	2.0	1.3	0.0	0.0	0.0
300	6.0	4.4	3.1	1.0	0.8	0.6	0.0	0.0	0.0
480	3.9	2.8	0.9	0.7	0.7	0.2	0.0	0.0	0.0
600	2.2	1.7	0.6	0.6	0.1	0.2	0.0	0.0	0.0

Notes: The figures are the percentage of all decisions at the stated significance level for which the test statistic rejects the model when the correct GRS statistic does not reject. (1) For Panel A, the sample periods cover July 1963 to December 2019, the sample window for the test is either 5, 10, 15, 20, 25, 40, or 50 years, the number of models tested are 6 for each window, and we average over 19 different sets of test assets, listed in Table A1. These windows overlap, adjusted in a rolling window, so that all but 1 year of data overlaps with the next sample window. This means that there are 53 samples for the 5-year window. The rejection rates are aggregated over the 6 models employed, so that a 54% value in the top leftmost cell corresponds to roughly 171 incorrect rejections on average. The factor models considered are the CAPM, the Fama–French 3 factor model, 4 and 5 factor models, as well as a six-factor model including momentum, all as considered in Fama and French (2016). (2) For Panel B, the sample periods cover January 1926 to December 2019. The models considered here number three, as factors for size, book-to-market, momentum and the market are all that are available. The test assets are constructed from industry classification, book-to-market crossed with size, momentum crossed with size, book-to-market, size and momentum decile portfolios. (3) For a detailed description of the factor and test asset construction see Fama and French (2015, 2016).

For both Panels A and B, we see a virtually monotonic decline in over-rejections as sample size increases, albeit at a fairly slow rate. The asymptotic χ^2 statistic over-rejects roughly 5% of the time relative to the $\tilde{W}$ statistic even with a 25-year window. The $\tilde{W}$ over-rejects roughly 5% of the time with 5 years of data, and over 1% of the time even with a 25-year window. The $\hat{W}$, which is fairly close to the correct $\tilde{W}$ statistic, does not appear to over-reject with more than 25 years of data, and over-rejects less than 0.1% of the time with 10 or more years of data. The simulations based on models and calibrations described in Appendix C.1 reveal that $\hat{W}$ always over-rejects, at least in this experimental design. The over-rejection declines with sample size, but at a decreasing (non-linear) rate, increases with the number of factors, and is largely unrelated to the number of test assets.

In Tables A5 and A6, we present the number of cases for which each statistic misranks factor models relative to the correct $\tilde{W}$ statistic's ranking by test value and p -value respectively. Misranking of models by the $\tilde{W}$ statistic is remarkably large, even at a 50-year data horizon if the set of models is as large as six, for either ranking method (statistic value or p -value). Replacing the correct form of the GRS test with $\tilde{W}$ scrambles rankings over 15% of the time even at the 50-year horizon, and over 50% of the time at the 5-year horizon, for cases with six models (Panel A). When there are only three models (Panel B), misrankings are naturally fewer, and for data horizons over 15 years misrankings occur mostly less than 5% of the time across methods. The $\hat{W}$ is typically consistent with the correct $\tilde{W}$ statistic, though misrankings occur even at the 40-year window length.

Table A5. Percentage of subsamples/models with different model rank outcomes from the correct GRS statistic $\tilde{W}$, ranked by test statistic value.

Window (Months)	Any Model Misranked		Top Model Misranked	
	$\tilde{W}$	$\hat{W}$	$\tilde{W}$	$\hat{W}$
Panel A: 1963–2019				
60	57.5	2.7	15.1	0.2
120	37.6	0.9	9.4	0.0
180	21.4	0.0	4.4	0.0
240	12.5	0.0	3.0	0.0
300	9.7	0.2	2.4	0.2
480	14.3	0.3	1.8	0.0
600	15.1	0.0	0.0	0.0
Panel B: 1926–2019				
60	17.4	0.7	9.6	0.4
120	10.6	0.2	5.5	0.0
180	5.2	0.2	1.9	0.0
240	3.1	0.0	0.9	0.0
300	1.9	0.0	1.2	0.0
480	1.5	0.3	0.6	0.0
600	1.1	0.0	0.7	0.0

Notes: The figures are the percentage of misrankings from a particular test statistic value across models relative to the GRS statistic ranking. (1) For Panel A, the sample periods cover July 1963 to December 2019, the sample window for the test is either 5, 10, 15, 20, 25, 40, or 50 years, the number of models tested is 6 for each window, and we aggregate over 19 different sets of test assets, listed in Table A1. These windows overlap, adjusted in a rolling window so that all but 1 year of data overlaps with the next sample window. The factor models considered are the CAPM, the Fama–French 3-factor model, 4 and 5-factor models, as well as a 6-factor model including momentum, all as considered in Fama and French (2016). (2) For Panel B, the sample periods cover January 1926 to December 2019. The models considered here number three, as factors for size, book-to-market, momentum and the market are all that are available. The test assets are constructed from industry classification, book-to-market crossed with size, momentum crossed with size, book-to-market, size and momentum decile portfolios. (3) For a detailed description of the factor and test asset construction see Fama and French (2015, 2016).

Table A6. Percentage of subsamples/models with different model rank outcomes from the correct GRS statistic $\tilde{W}$, ranked by test p -value.

Window (Months)	Any Model Mis-Ranked		Top Model Mis-Ranked	
	$\tilde{W}$	$\hat{W}$	$\tilde{W}$	$\hat{W}$
Panel A: 1963–2019				
60	54.6	1.9	13.1	0.4
120	37.4	0.5	9.1	0.0
180	21.9	0.4	4.4	0.1
240	11.5	0.1	2.8	0.0
300	9.4	0.2	2.1	0.0
480	13.2	0.3	1.8	0.0
600	13.2	0.0	0.0	0.0
Panel B: 1926–2019				
60	15.0	0.7	8.1	0.7
120	10.2	0.0	5.5	0.0
180	5.4	0.0	2.1	0.0
240	3.3	0.2	0.9	0.0
300	1.9	0.0	1.2	0.0
480	1.2	0.0	0.6	0.0
600	1.1	0.0	0.7	0.0

Notes: The figures are the percentage of misrankings from a particular test statistic p -value across models relative to the GRS statistic ranking. (1) For Panel A the sample periods cover July 1963 to December 2019, the sample window for the test is either 5, 10, 15, 20, 25, 40, or 50 years, the number of models tested is 6 for each window, and we aggregate over 19 different sets of test assets, listed in Table A1. These windows overlap, adjusted in a rolling window so that all but 1 year of data overlaps with the next sample window. The factor models considered are the CAPM, the Fama–French 3-factor model, 4 and 5-factor models, as well as a 6-factor model including momentum, all as considered in Fama and French (2016). (2) For Panel B, the sample periods cover January 1926 to December 2019. The models considered here number three, as factors for size, book-to-market, momentum and the market are all that are available. The test assets are constructed from industry classification, book-to-market crossed with size, momentum crossed with size, book-to-market, size and momentum decile portfolios. (3) For a detailed description of the factor and test asset construction see Fama and French (2015, 2016).

If we restrict our attention to cases for which the top model is misranked, the $\hat{W}$ statistic is consistent with the correct $\tilde{W}$ statistic once we have 40 or more years of data, but $\tilde{W}$ misranks even at 40 years of data. Again, when there are only three models (Panel B), misrankings are fewer, and for data horizons over 10 years misrankings occur less than 5% of the time across methods.

Finally, in Table A7, we present evidence on rankings from these different test statistics compared to rankings based on the p -values of these test statistics, to see if they are consistent. We can think of ranking by the test statistic as a sort of mean squared error or model R^2 ranking - perhaps helpful if we are interested in minimizing model prediction error even if models are false (see, for instance, Teräsvirta and Mellin 1986). Here, we see that all the rankings by test statistics are fragile, even at a 50 year horizon, averaging close to 10% misranked if we compare three or six asset pricing models to each other. Table A7 highlights that the statistically sound p -values, instead of the raw GRS statistics, should be used in model ranking.

Considering the case of a true model that includes a subset of the available factors, an untabulated analysis of the simulated test rankings formed using the magnitude of the GRS statistic confirms that the probability of an incorrect model, larger than the true model, achieving a high rank versus other models increases with the number of factors in the model, when the GRS statistic ranking differs from the p -value ranking. This bias is stronger when using an incorrect GRS statistic. In cases for which both the GRS statistic ranking and the p -value ranking agree, there is no such pattern to tilt to larger models than the true model.

The important insight to take away from these results is that the error in calculating the GRS statistic can have a material impact on empirical results, particularly when twenty

or fewer years of data are used, which is not uncommon in empirical asset pricing studies. For example, Barillas and Shanken (2018) performed model comparisons on a little less than 15 years of monthly data. Harvey and Liu (2021) considered tests of asset pricing models and report simulations for 20 and 40 years of monthly data. Sha and Gao (2019) used 144 months of data and exploited 6 metrics to evaluate factor model performance, including the GRS statistic. Baek and Bilson (2015) considered 234 months of data in a subsample estimation. Chiah et al. (2016) used 23 years of data when comparing models using the GRS statistic. One takeaway from these papers is that many situations involving specialized data (like Sha and Gao 2019, exploring mutual fund returns in China) or subsample robustness checks (like Baek and Bilson 2015) are necessarily constrained to shorter samples than fifty or even twenty years, so that the bias from an incorrectly calculated GRS statistic becomes large.

Table A7. Percentage of subsamples/models with different model rankings if ranked by *p*-values rather than test statistics.

Window (Months)	$\widehat{W}$	$\widetilde{W}$	$\widehat{W}$
Panel A: 1963–2019			
60	16.3	21.8	16.8
120	4.9	5.8	5.3
180	4.3	3.9	3.9
240	4.6	5.4	4.4
300	4.8	4.8	4.8
480	2.6	3.8	2.6
600	6.6	7.9	6.6
Panel B: 1926–2019			
60	2.2	5.0	2.6
120	0.6	1.2	0.8
180	0.4	0.2	0.6
240	0.4	0.2	0.2
300	0.7	0.7	0.7
480	5.5	5.8	5.8
600	7.4	7.4	7.4

Notes: The figures are the percentage of misrankings from a particular test statistic value across models relative to the GRS statistic ranking. (1) For Panel A, the sample periods cover July 1963 to December 2019, the sample window for the test is either 5, 10, 15, 20, 25, 40, or 50 years; the number of models tested are 6 for each window; and we average over 19 different sets of test assets, listed in Table A1. These windows overlap, adjusted in a rolling window so that all but 1 year of data overlaps with the next sample window. The factor models considered are the CAPM, the Fama–French 3-factor model, 4 and 5-factor models, as well as a 6-factor model including momentum, all as considered in Fama and French (2016). (2) For Panel B, the sample periods cover January 1926 to December 2019. The models considered here number three, as factors for size, book-to-market, momentum and the market are all that are available. The test assets are constructed from industry classification, book-to-market crossed with size, momentum crossed with size, book-to-market, size and momentum decile portfolios. (3) For a detailed description of the factor and test asset construction see Fama and French (2015, 2016).

We do not recommend ranking models with the magnitude of the GRS statistic, and instead suggest the use of the *p*-value of the statistic from the exact F-distribution, since the *p*-value internalizes the different degrees of freedom of the GRS statistics computed for models with a different number of factors. We recognize that ranking of models by the GRS statistic has a desirable economic intuition—this is a direct, easy-to-understand metric tied to a model’s factors spanning the test asset returns. But researchers need to at least understand that this ranking might have undesirable statistical properties. The detailed results that these tables are based on, and additional summary statistics are available on request.

Appendix C. Simulation Results

Here, we present simulation results to evaluate the size performance of the alternative tests $\tilde{W}$, $\hat{W}$, $\check{W}$ and χ^2 . We conducted three sets of simulations, differentiated by the error generating process. Within each of these three sets of simulations, we look at two different factor models, one with three factors and one with six factors, and we look at two different groupings of test assets, one grouping being decile portfolios and one being a five by five set of twenty five portfolios. For each of these, we consider monthly samples of lengths 5, 10, 15, 20, 25, 40, and 50 years.

The first two sets of simulation results exploit the normal and t-distributions for the errors, calibrating return moments (mean, volatility, covariance) to French’s portfolio returns. For the decile returns, we calibrate to the size-sorted decile monthly returns over 1963/7–2019/12, and for the 5 by 5 test asset case, we calibrate to the size by book-to-market returns, again over 1963/7–2019/12. The last set of simulations employed bootstrap methods and the French portfolio returns, described fully below.

Appendix C.1. Simulated Normal and t-Distributed Returns

Following the literature, we generate portfolio excess returns $\tilde{r}_{pt}$ as normal, independent and identically distributed, calibrated to monthly U.S. stock returns. We also consider the t-distribution with eight degrees of freedom and a bootstrap simulation.

The excess return for test asset i and time t is generated based on model (1), which is rewritten here:

$$\tilde{r}_{it} = \delta_{i0} + \sum_{j=1}^L \delta_{ij}\tilde{r}_{jt} + \tilde{\eta}_{it}, \tag{A11}$$

where $\tilde{\eta}_{it} \sim iid$ normal across t with mean 0 and volatility σ_{ii} , and $\tilde{r}_{jt} \sim iid$ normal across t with mean (μ_j/L) , volatility σ_j and $E[\tilde{\eta}_{it}\tilde{r}_{jt}] = 0$. We set $\mu_j = 0.01$, $\sigma_j = 0.02$, $\sigma_{ii} = 0.08$, $\delta_{ij} = 1, \forall i, j$ and we explore only the case of $\delta_{i0} = 0, \forall i$.

We explore the size properties of the correct and incorrect formulas of the GRS statistic for numbers of portfolios (L) from 3 to 6, test assets (N) from 10 to 25, and sample sizes (T) from 60 (months) to 600. This spans typical applications of the GRS statistic. Our simulations show that the performance of the incorrect GRS formula generally suffers deterioration as the number of firms and factors increases, as one might expect. We present simulation results for the normal case in Tables A8 and A9 for the t-distribution case.

Table A8. Null rejection rates.

Test Assets/Portfolios N/L	Years	$\tilde{W}$		$\hat{W}$		$\check{W}$		χ^2	
		1%	10%	1%	10%	1%	10%	1%	10%
10 / 3	5	0.010	0.104	0.011	0.106	0.016	0.132	0.062	0.240
	10	0.011	0.100	0.011	0.102	0.013	0.119	0.026	0.162
	15	0.010	0.099	0.010	0.100	0.012	0.109	0.020	0.135
	20	0.008	0.094	0.008	0.096	0.010	0.101	0.016	0.120
	25	0.010	0.100	0.010	0.101	0.011	0.107	0.017	0.122
	40	0.009	0.102	0.009	0.103	0.010	0.106	0.012	0.113
	50	0.009	0.096	0.009	0.096	0.009	0.099	0.012	0.105
10 / 6	5	0.009	0.104	0.010	0.110	0.021	0.158	0.065	0.246
	10	0.010	0.095	0.010	0.097	0.015	0.124	0.026	0.156
	15	0.011	0.105	0.011	0.107	0.014	0.125	0.020	0.143
	20	0.009	0.094	0.009	0.096	0.011	0.111	0.015	0.124
	25	0.010	0.103	0.010	0.104	0.012	0.114	0.015	0.125
	40	0.009	0.097	0.010	0.098	0.011	0.104	0.012	0.111
	50	0.009	0.098	0.009	0.099	0.010	0.103	0.011	0.109

Table A8. Cont.

Test Assets/Portfolios N/L	Years	$\widetilde{W}$		$\widehat{W}$		$\check{W}$		χ^2	
		1%	10%	1%	10%	1%	10%	1%	10%
25 / 3	5	0.011	0.103	0.011	0.107	0.016	0.141	0.484	0.730
	10	0.008	0.101	0.008	0.104	0.011	0.122	0.124	0.367
	15	0.010	0.099	0.010	0.101	0.012	0.115	0.062	0.260
	20	0.009	0.099	0.010	0.101	0.011	0.110	0.043	0.206
	25	0.010	0.097	0.011	0.098	0.012	0.106	0.034	0.180
	40	0.009	0.100	0.009	0.101	0.009	0.105	0.020	0.147
	50	0.008	0.093	0.008	0.094	0.009	0.098	0.018	0.136
25 / 6	5	0.008	0.097	0.009	0.104	0.020	0.163	0.524	0.763
	10	0.010	0.102	0.011	0.106	0.018	0.140	0.130	0.378
	15	0.011	0.103	0.012	0.107	0.016	0.131	0.069	0.261
	20	0.012	0.103	0.012	0.105	0.016	0.124	0.046	0.212
	25	0.009	0.100	0.010	0.102	0.013	0.116	0.031	0.181
	40	0.009	0.099	0.009	0.100	0.011	0.111	0.020	0.149
	50	0.011	0.099	0.011	0.100	0.012	0.107	0.019	0.135

Notes: (1) Bold-faced numbers are rejection rates three standard deviations larger than the nominal values. (2) These results are based on 10,000 simulations. (3) The models are $\bar{r}_{it} = \delta_{i0} + \sum_{j=1}^L \delta_{ij} \bar{r}_{jt} + \bar{\eta}_{it}$, where $\bar{r}_{jt} \sim iid \mathcal{N}(\mu_j/L, \sigma_j^2)$, $\bar{\eta}_{it} \sim iid \mathcal{N}(0, \sigma_{ii}^2)$ and $E[\bar{\eta}_{it} \bar{r}_{jt}] = 0$. For $\forall j$, $\mu_j = 0.01$, $\sigma_j = 0.02$, $\sigma_{ii} = 0.08$, and $\delta_{ij} = 1$.

Table A9. Null rejection rates.

Test Assets/Portfolios N/L	Years	$\widetilde{W}$		$\widehat{W}$		$\check{W}$		χ^2	
		1%	10%	1%	10%	1%	10%	1%	10%
10 / 3	5	0.009	0.104	0.009	0.107	0.015	0.137	0.061	0.236
	10	0.011	0.098	0.011	0.100	0.012	0.113	0.026	0.158
	15	0.009	0.100	0.009	0.101	0.011	0.109	0.018	0.135
	20	0.010	0.100	0.010	0.101	0.011	0.107	0.017	0.125
	25	0.010	0.099	0.010	0.099	0.011	0.106	0.014	0.122
	40	0.010	0.106	0.010	0.106	0.010	0.109	0.013	0.120
	50	0.009	0.100	0.009	0.101	0.010	0.105	0.012	0.112
10 / 6	5	0.010	0.104	0.012	0.110	0.023	0.163	0.068	0.256
	10	0.010	0.095	0.010	0.098	0.015	0.122	0.027	0.155
	15	0.010	0.099	0.011	0.101	0.014	0.117	0.020	0.136
	20	0.009	0.096	0.009	0.098	0.012	0.111	0.016	0.123
	25	0.012	0.097	0.012	0.098	0.014	0.109	0.017	0.120
	40	0.010	0.099	0.010	0.099	0.011	0.105	0.014	0.111
	50	0.010	0.102	0.011	0.103	0.011	0.109	0.012	0.114
25 / 3	5	0.009	0.095	0.009	0.098	0.014	0.131	0.477	0.725
	10	0.011	0.105	0.011	0.107	0.014	0.127	0.129	0.370
	15	0.011	0.103	0.011	0.105	0.013	0.117	0.064	0.254
	20	0.011	0.094	0.011	0.096	0.012	0.106	0.041	0.206
	25	0.009	0.095	0.009	0.096	0.010	0.104	0.031	0.180
	40	0.009	0.098	0.009	0.099	0.010	0.104	0.021	0.145
	50	0.010	0.097	0.010	0.098	0.011	0.102	0.017	0.137
25 / 6	5	0.009	0.100	0.010	0.106	0.022	0.169	0.533	0.772
	10	0.010	0.103	0.011	0.108	0.020	0.140	0.131	0.380
	15	0.011	0.104	0.012	0.108	0.015	0.132	0.069	0.266
	20	0.010	0.101	0.010	0.103	0.015	0.117	0.044	0.209
	25	0.010	0.099	0.010	0.100	0.012	0.114	0.032	0.180
	40	0.009	0.102	0.010	0.104	0.011	0.114	0.021	0.150
	50	0.009	0.101	0.010	0.101	0.011	0.109	0.019	0.141

Notes: (1) Bold-faced numbers are rejection rates three standard deviations larger than the nominal values. (2) The results are based on 10,000 simulations. (3) The models are $\bar{r}_{it} = \delta_{i0} + \sum_{j=1}^L \delta_{ij} \bar{r}_{jt} + \bar{\eta}_{it}$, where $\bar{r}_{jt} \sim iid (\mu_j/L, \sigma_j^2)$, $\bar{\eta}_{it} \sim iid (0, \sigma_{ii}^2)$, t with 8 df, and $E[\bar{\eta}_{it} \bar{r}_{jt}] = 0$. For $\forall j$, $\mu_j = 0.01$, $\sigma_j = 0.02$, $\sigma_{ii} = 0.08$, and $\delta_{ij} = 1$.

The correct formula of the GRS statistic generally presents no evidence of incorrect size in Table A8, as our simulation setting is one in which it should have the correct small-sample exact size. Even with the t-distribution, the GRS performs well. The $\widehat{W}$ formula shows some evidence of over-rejection with the t-distribution and a small sample size. The $\widehat{W}$ formula of the GRS statistic and the χ^2 show strong over-rejection under 20 years of data and the χ^2 shows evidence of over-rejection even with 50 years of data.

Table A10. Null rejection rates.

Test Assets/Portfolios N/L	Years	$\widehat{W}$		$\widehat{W}$		$\widehat{W}$		χ^2	
		1%	10%	1%	10%	1%	10%	1%	10%
10 / 3	5	0.097	0.290	0.097	0.292	0.115	0.333	0.228	0.458
	10	0.140	0.344	0.140	0.344	0.153	0.367	0.204	0.425
	15	0.167	0.372	0.167	0.372	0.176	0.385	0.210	0.417
	20	0.189	0.395	0.189	0.395	0.198	0.405	0.222	0.433
	25	0.202	0.408	0.202	0.408	0.207	0.417	0.230	0.436
	40	0.234	0.431	0.234	0.431	0.237	0.435	0.250	0.447
	50	0.249	0.450	0.249	0.450	0.251	0.455	0.261	0.465
10 / 6	5	0.080	0.266	0.080	0.268	0.114	0.342	0.213	0.451
	10	0.124	0.323	0.124	0.324	0.146	0.361	0.183	0.402
	15	0.148	0.356	0.149	0.357	0.164	0.381	0.192	0.406
	20	0.173	0.374	0.173	0.374	0.184	0.394	0.203	0.415
	25	0.186	0.387	0.186	0.387	0.194	0.403	0.208	0.418
	40	0.213	0.417	0.213	0.417	0.218	0.425	0.227	0.434
	50	0.234	0.437	0.234	0.437	0.239	0.444	0.246	0.450
25 / 3	5	0.021	0.144	0.021	0.145	0.031	0.184	0.546	0.776
	10	0.023	0.156	0.024	0.157	0.031	0.182	0.185	0.452
	15	0.026	0.167	0.026	0.167	0.031	0.183	0.113	0.349
	20	0.032	0.175	0.032	0.176	0.036	0.188	0.091	0.304
	25	0.037	0.177	0.037	0.177	0.039	0.188	0.082	0.279
	40	0.042	0.188	0.042	0.188	0.044	0.196	0.069	0.248
	50	0.044	0.192	0.044	0.192	0.047	0.197	0.067	0.235
25 / 6	5	0.018	0.132	0.018	0.134	0.034	0.208	0.583	0.793
	10	0.022	0.145	0.023	0.146	0.032	0.194	0.185	0.449
	15	0.023	0.150	0.023	0.151	0.033	0.183	0.103	0.329
	20	0.025	0.158	0.025	0.158	0.032	0.182	0.078	0.288
	25	0.027	0.160	0.027	0.161	0.031	0.179	0.067	0.255
	40	0.030	0.167	0.030	0.167	0.033	0.181	0.056	0.228
	50	0.034	0.167	0.034	0.167	0.036	0.176	0.052	0.209

Notes: (1) Bold-faced numbers are rejection rates three standard deviations larger than the nominal values. (2) The results are based on 10,000 simulations. (3) The models are $\hat{r}_{it} = \delta_{i0} + \sum_{j=1}^L \delta_{ij} \hat{r}_{jt} + \hat{\eta}_{it}$, where the data were generated through a block-bootstrap approach.

Appendix C.2. Bootstrap Simulation

There are two main categories of bootstrapping in the regression context, the random X approach, which resamples the complete set of variables including the dependent variable for each observation, and the fixed X approach, which resamples residuals and explanatory variable values and forms simulated dependent variable values. That is, the fixed X approach builds simulated dependent variable values from the explanatory variables and either simulated or resampled regression residuals. The choice of using either simulated or resampled residuals is what distinguishes the major variations of the fixed X bootstrap approach. The non-parametric fixed X bootstrap approach, which we employ, uses resampled regression residuals.

Suppose we have a sample of T observations of a dependent variable $r_{i,t}$, ($i = 1, \dots, N$), a $K \times 1$ vector of factor portfolios $r_{p,t}$, and a regression model $E[r_{i,t}] = \alpha_i + \beta_i' r_{p,t}$. Define $\hat{\alpha}_i$ and $\hat{\beta}_i$ as the OLS estimates of α_i and β_i and, noting that we wish to explore the null

hypothesis that $\alpha_i = 0$, define $\hat{r}_{i,t}^* = \hat{\beta}_i r_{p,t}$ and $\hat{\varepsilon}_{i,t} = r_{i,t} - (\hat{\alpha}_i + \hat{\beta}_i r_{p,t})$. We then form R resamples of $\hat{r}_{i,t}^*$, ($i = 1, \dots, N$), and $r_{p,t}$ with each resample containing T observations. Separately and independently, we form R resamples of $\hat{\varepsilon}_{i,t}$ with each resample also containing T observations, and finally we form $r_{i,t}^* = \hat{r}_{i,t}^* + \hat{\varepsilon}_{i,t}$ for each of the R resamples. Using $r_{i,t}^*$ and $r_{p,t}$, we fit the model $E[r_{i,t}^*] = \alpha_i + \beta_i r_{p,t}$ on each of these resampled datasets, and retrieve the various GRS test statistics for each resampled dataset.

To deal with a well-documented property of financial returns, lack of independence across time, we also employ block bootstrap resampling which allows for data dependence. See, for instance, Politis and Romano (1994), White (2000), and Gonçalves and White (2002, 2005). It is the resampling in (random-length) blocks from the original data that produces results incorporating data dependence. Politis and Romano (1994) used blocks of data with lengths distributed according to the geometric distribution. The mean block length b is data-dependent. Politis and Romano (1994) recommended a length proportional to $T^{1/3}$, where T = sample size, which is what we use.

Again, we exploit French's portfolio returns. For the decile returns, we resample from the size-sorted decile monthly returns over 1963/7–2019/12, and for the 5 by 5 test asset case we resample from the size by book-to-market returns, over 1963/7–2019/12. Bootstrap simulations show persistent over-rejection of the null hypothesis in all these tests, though the correct GRS F test shows the smallest over-rejection. Similarly to Harvey and Liu (2021), we find little or no over-rejection when we evaluate t -tests on the intercept with bootstrapped data, rather than a joint test across intercepts of the test assets, but joint tests appear much more fragile than the one-at-a-time t -tests on intercepts.

Appendix D. Software Packages

The SAS and R packages used to implement our generalized GRS test can be found at the authors' websites: <http://markkamstra.com/data.html> (accessed on 31 August 2023) (SAS) and <https://ruoyaoshi.github.io/> (accessed on 31 August 2023) (R). A Stata package `grsftest` coded by Mengnan (Cliff) Zhu can be found at <https://ideas.repec.org/c/boc/bocode/s458828.html> (accessed on 31 August 2023). See Zhu (2020).

Appendix E. Detailed Model Statistics

Appendix E.1. 5 Year Windows over 1963–2019

Appendix E.2. 10 Year Windows over 1963–2019

Appendix E.3. 15 Year Windows over 1963–2019

Appendix E.4. 20 Year Windows over 1963–2019

Appendix E.5. 25 Year Windows over 1963–2019

Appendix E.6. 40 Year Windows over 1963–2019

Appendix E.7. 50 Year Windows over 1963–2019

Appendix E.8. 5 Year Windows over 1926–2019

Appendix E.9. 10 Year Windows over 1926–2019

Appendix E.10. 15 Year Windows over 1926–2019

Appendix E.11. 20 Year Windows over 1926–2019

Appendix E.12. 25 Year Windows over 1926–2019

Appendix E.13. 40 Year Windows over 1926–2019

Appendix E.14. 50 Year Windows over 1926–2019

Available on Supplementary Materials.

Appendix F. Summary Statistics

Appendix F.1. 5 Year Windows over 1963–2019

Appendix F.2. 10 Year Windows over 1963–2019

Appendix F.3. 15 Year Windows over 1963–2019

Appendix F.4. 20 Year Windows over 1963–2019

Appendix F.5. 25 Year Windows over 1963–2019

Appendix F.6. 40 Year Windows over 1963–2019

Appendix F.7. 50 Year Windows over 1963–2019

Appendix F.8. 5 Year Windows over 1926–2019

Appendix F.9. 10 Year Windows over 1926–2019

Appendix F.10. 15 Year Windows over 1926–2019

Appendix F.11. 20 Year Windows over 1926–2019

Appendix F.12. 25 Year Windows over 1926–2019

Appendix F.13. 40 Year Windows over 1926–2019

Appendix F.14. 50 Year Windows over 1926–2019

Available on Supplementary Materials.

Notes

- ¹ See, for instance, Cakici et al. (2013, eq. (4)) and Mosoeu and Kodongo (2022, eq. (2)).
- ² Asymptotic versions are commonly employed or promoted. See, for instance, MacKinlay and Richardson (1991), Cochrane (2005, p. 234), Zaremba and Czapkiewicz (2017), Demaj et al. (2018), Belimam et al. (2018), Qin (2019), and Verbeek (2021, Sect. 2.3).
- ³ We acknowledge that it is difficult to take seriously the assumption of normality of returns—returns are bounded below by -100% due to limited liability in financial markets for publicly traded assets and returns are known to be heteroskedastic and dependent over time. Here we adopt the Gibbons et al. (1989) setting for comparison purposes and to develop small sample results. Knez and Ready (1997) develop some interesting approaches for factor model estimation allowing for non-normality.
- ⁴ For related analysis on an extension to the GRS test, see Kleibergen and Zhan (2020) and Kleibergen et al. (2023).
- ⁵ Cochrane (2005, eq. (12.6)) uses $\tilde{\Omega}$ for Ω and $\tilde{\Sigma}$ for Σ , but pre-multiplies by $\frac{T-N-L}{N}$, so that the resulting GRS statistic equals to $\tilde{W}$ in this paper. The d.f. adjustment (or lack of it) in the estimators of Σ can be easily offset by pre-multiplying an appropriate factor, but this is not the case for Ω .
- ⁶ Perhaps another outcome of Cochrane (2005), the Stata packages pre-multiply the ratio $\frac{T-N-L}{N}$ used by Cochrane (2005), but fail to use the corresponding $\tilde{\Sigma}$.
- ⁷ We need to point out that the argument here is based on an approximation, as $E(\tilde{r}_p' \tilde{\Omega}^{-1} \tilde{r}_p)$ is a non-linear function of $\tilde{r}_p$ and $\tilde{\Omega}$. Moreover, $\tilde{r}_p' \tilde{\Omega}^{-1} \tilde{r}_p$ may deviate from its mean for a particular sample. Therefore, it is entirely possible that the incorrect formula of the GRS statistic favors larger models in some cases.
- ⁸ See Tharyan (2009) and Ibert (2014).
- ⁹ For comparison, the generalized GRS statistic formula given in Cochrane (2005, p. 230) uses $\tilde{\Omega}$ and $\tilde{\Sigma}$, but Cochrane (2005) is careful to properly pre-multiply the ratio $\frac{T-N-L}{N}$ so that the statistic equals $\tilde{W}$ in this paper.
- ¹⁰ This is because the F_{d_1, d_2} and the $\chi^2_{d_1}$ distributions are related through $d_1 F_{d_1, d_2} \xrightarrow{d} \chi^2_{d_1}$ as $d_2 \rightarrow \infty$, where d_1 and d_2 are the degrees of freedom.
- ¹¹ Several different statistics are all commonly used in empirical research. For example, the usual Wald statistic equals $T(1 + \tilde{r}_p' \tilde{\Omega}^{-1} \tilde{r}_p)^{-1} \tilde{\delta}_0' \tilde{\Sigma}^{-1} \tilde{\delta}_0$, which deviates from $N\tilde{W}$ by a factor of $\frac{T-L-1}{T-N-L}$. Another example is the χ^2 statistic formula in Cochrane (2005, p. 230), which deviates from $N\tilde{W}$ by a factor of $\frac{T}{T-N-L}$. Both factors are larger than 1, meaning that both χ^2 statistics are larger than $N\tilde{W}$ for any sample size, with Cochrane's (2005, p. 230) formula being the largest. Note that Cochrane (2005, p. 230), like Gibbons et al. (1989), only gives the formula for the $L = 1$ case, and here we refer to its generalized version for the $L \geq 1$ case.
- ¹² In Section 3, we only report empirical results using the Wald statistic for the asymptotic χ^2 test. Since the formula in Cochrane (2005, p. 230) is even larger, it will lead to worse over-rejection.
- ¹³ We thank Ken French for making this valuable resource freely available.

References

- Affleck-Graves, John, and Bill McDonald. 1989. Nonnormalities and tests of asset pricing theories. *The Journal of Finance* 44: 889–908. [CrossRef]
- Alhomaidi, Asem, M. Kabir Hassan, William J. Hippler, and Abdullah Mamun. 2019. The impact of religious certification on market segmentation and investor recognition. *Journal of Corporate Finance* 55: 28–48. [CrossRef]
- Alshammari, Saad, and Shingo Goto. 2022. What factors drive Saudi stock markets? Firm characteristics that attract retail trades. *International Review of Economics and Finance* 80: 994–1011. [CrossRef]
- Anderson, Theodore Wilbur. 2003. *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Hoboken: Wiley-Interscience.
- Baek, Seungho, and John F. O. Bilson. 2015. Size and value risk in financial firms. *Journal of Banking and Finance* 55: 295–326. [CrossRef]
- Barillas, Francisco, and Jay Shanken. 2018. Comparing asset pricing models. *The Journal of Finance* 73: 715–54. [CrossRef]
- Bartlett, Maurice. 1951. An Inverse Matrix Adjustment Arising in Discriminant Analysis. *The Annals of Mathematical Statistics* 22: 107–11. [CrossRef]
- Belimam, Doha, Yong Tan, and Ghizlane Lakhnati. 2018. An empirical comparison of asset-pricing models in the Shanghai a-share exchange market. *Asia-Pacific Financial Markets* 25: 249–65. [CrossRef]
- Cakici, Nusret, Frank J. Fabozzi, and Sinan Tan. 2013. Size, value, and momentum in emerging market stock returns. *Emerging Markets Review* 16: 46–65. [CrossRef]
- Chiah, Mardy, Daniel Chai, Angel Zhong, and Song Li. 2016. A Better Model? An empirical investigation of the Fama-French five-factor model in Australia. *International Review of Finance* 16: 595–638. [CrossRef]
- Choi, Seo Joon, Kanghyun Kim, and Sunyoung Park. 2020. Is systemic risk systematic? Evidence from the US stock markets. *International Journal of Finance and Economics* 25: 642–63. [CrossRef]
- Cochrane, John. 2005. *Asset Pricing: Revised Edition*. Princeton: Princeton University Press.
- Demaj, Arber, Bora Oskay, Betim Lushtaku, and Thedo Linssen. 2018. Asset Pricing Models and Anomalies: An Empirical Analysis. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3128562 (accessed on 4 April 2022).
- Fama, Eugene F., and Kenneth R. French. 2015. A Five-Factor Asset Pricing Model. *Journal of Financial Economics* 116: 1–22. [CrossRef]
- Fama, Eugene F., and Kenneth R. French. 2016. Dissecting Anomalies with a Five-Factor Model. *The Review of Financial Studies* 29: 69–103. [CrossRef]
- Fama, Eugene F., and Kenneth R. French. 2018. Choosing factors. *Journal of Financial Economics* 128: 234–52. [CrossRef]
- Ferson, Wayne E., and Stephen R. Foerster. 1994. Finite sample properties of the generalized method of moments in tests of conditional asset pricing models. *Journal of Financial Economics* 36: 29–55. [CrossRef]
- Gibbons, Michael R., Stephen A. Ross, and Jay Shanken. 1989. A Test of the Efficiency of a Given Portfolio. *Econometrica* 57: 1121–52. [CrossRef]
- Gonçalves, Silvia, and Halbert White. 2002. The Bootstrap of the Mean for Dependent Heterogenous Arrays. *Econometric Theory* 18: 1367–84. [CrossRef]
- Gonçalves, Silvia, and Halbert White. 2005. Bootstrap Standard Error Estimates for Linear Regressions. *Journal of the American Statistical Association* 100: 970–79. [CrossRef]
- Hanauer, Matthias X. 2020. A Comparison of Global Factor Models. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3546295 (accessed on 19 April 2022).
- Harvey, Campbell R., and Yan Liu. 2021. Lucky factors. *Journal of Financial Economics* 141: 413–35. [CrossRef]
- Hayashi, Fumio. 2000. *Econometrics*. Princeton: Princeton University Press.
- Ibert, Markus. 2014. *GRSTEST2: Stata Module to Implement the Gibbons, Ross, Shanken (1989) Test*. Statistical Software Components S457786. Boston: Boston College Department of Economics.
- Kan, Raymond, Xiaolu Wang, and Xinghua Zheng. 2024. In-sample and out-of-sample Sharpe ratios of multi-factor asset pricing models. *Journal of Financial Economics* 155: 103837. [CrossRef]
- Kim, Jae H. 2022. GRS.test: GRS Test for Portfolio Efficiency, Its Statistical Power Analysis, and Optimal Significance Level Calculation. R Package Version 1.2. Available online: <https://CRAN.R-project.org/package=GRS.test> (accessed on 1 April 2022).
- Kleibergen, Frank, and Zhaoguo Zhan. 2020. Robust inference for consumption-based asset pricing. *The Journal of Finance* 75: 507–50. [CrossRef]
- Kleibergen, Frank, Lingwei Kong, and Zhaoguo Zhan. 2023. Identification robust testing of risk premia in finite samples. *Journal of Financial Econometrics* 21: 263–97. [CrossRef]
- Knez, Peter J., and Mark J. Ready. 1997. On the robustness of size and book-to-market in cross-sectional regressions. *The Journal of Finance* 52: 1355–82.
- Kroencke, Tim A. 2017. Asset pricing without garbage. *The Journal of Finance* 72: 47–98. [CrossRef]
- Leite, André Luis, Marcelo Cabus Klotzle, Antonio Carlos Figueiredo Pinto, and Aldo Ferreira da Silva. 2018. Size, value, profitability, and investment: Evidence from emerging markets. *Emerging Markets Review* 36: 45–59. [CrossRef]
- Lewellen, Jonathan, Stefan Nagel, and Jay Shanken. 2010. A skeptical appraisal of asset pricing tests. *Journal of Financial Economics* 96: 175–94. [CrossRef]
- MacKinlay, A. Craig, and Matthew P. Richardson. 1991. Using generalized method of moments to test mean-variance efficiency. *The Journal of Finance* 46: 511–27.

- Merdad, Hesham Jamil, M. Kabir Hassan, and William J. Hippler, III. 2015. The Islamic risk factor in expected stock returns: An empirical study in Saudi Arabia. *Pacific-Basin Finance Journal* 34: 293–314. [CrossRef]
- Mosoeu, Selebogo, and Odongo Kodongo. 2022. The Fama-French five-factor model and emerging market equity returns. *The Quarterly Review of Economics and Finance* 85: 55–76. [CrossRef]
- Politis, Dimitris N., and Joseph P. Romano. 1994. The Stationary Bootstrap. *Journal of the American Statistical Association* 89: 1303–13. [CrossRef]
- Qin, Rui. 2019. Study on Applicability of Fama-French Five-Factor Model in Chinese A-Share Market. Paper presented at the 2nd International Symposium on Social Science and Management Innovation (SSMI 2019), Xi'an, China, November 29–30; Amsterdam: Atlantis Press, pp. 491–500.
- Rouwenhorst, K. Geert. 1999. Local return factors and turnover in emerging stock markets. *The Journal of Finance* 54: 1439–64. [CrossRef]
- Sha, Yezhou, and Ran Gao. 2019. Which is the best: A comparison of asset pricing factor models in Chinese mutual fund industry. *Economic Modelling* 83: 8–16. [CrossRef]
- Sharpe, William F. 1964. Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. *Journal of Finance* 19: 425–42.
- Sharpe, William F. 1966. Mutual fund performance. *The Journal of Business* 39: 119–38. [CrossRef]
- Teräsvirta, Timo, and Ilkka Mellin. 1986. Model selection criteria and model selection tests in regression models. *Scandinavian Journal of Statistics*, 159–71.
- Tharyan, Rajesh. 2014. *GRSTEST: Stata Module to Implement the Gibbons et al. (1989) Test in a Single-Factor or Multi-Factor Setting*. Statistical Software Components S457069. Boston: Boston College Department of Economics.
- Verbeek, Marno. 2021. Panel Methods for Finance. In *Panel Methods for Finance*. Berlin: De Gruyter.
- White, H. 2000. A Reality Check for Data Snooping. *Econometrica* 68: 1097–26.
- Zaremba, Adam, and Anna Czapkiewicz. 2017. Digesting anomalies in emerging European markets: A comparison of factor pricing models. *Emerging Markets Review* 31: 1–15. [CrossRef]
- Zhu, Mengnan. 2020. *GRSFTEST: Stata Module to Perform the Gibbons, Ross, Shanken Test of Mean-Variance Efficiency of Asset Returns*. Statistical Software Components S458828. Boston: Boston College Department of Economics.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

Downside Risk in Australian and Japanese Stock Markets: Evidence Based on the Expectile Regression

Kohei Marumo ^{1,*} and Steven Li ²

¹ Faculty of Economics, Saitama University, 255 Shimo-Okubo, Sakura, Saitama 338-8570, Japan

² Graduate School of Business and Law (GSBL), RMIT University Australia, Melbourne 3000, Australia; steven.li@rmit.edu.au

* Correspondence: marumo@mail.saitama-u.ac.jp

Abstract: The expectile-based Value at Risk (EVaR) has gained popularity as it is more sensitive to the magnitude of extreme losses than the conventional quantile-based VaR (QVaR). This paper applies the expectile regression approach to evaluate the EVaR of stock market indices of Australia and Japan. We use an expectile regression model that considers lagged returns and common risk factors to calculate the EVaR for each stock market and to evaluate the interdependence of downside risk between the two markets. Our findings suggest that both Australian and Japanese stock markets are affected by their past development and the international stock markets. Additionally, ASX 200 index has significant impact on Nikkei 225 in terms of downside tail risk, while the impact of Nikkei 225 on ASX is not significant.

Keywords: downside risk; expectile regression; value at risk (EVaR); expected shortfall (ES); stock market

1. Introduction

Globalization and rapid technological advancements have led to an increasing integration of financial markets worldwide. As a result, the linkages and interdependence between national stock markets may have strengthened, which can have significant implications for corporate investment and financing strategies and international diversification. Strong interdependence would reduce the potential gains from international diversifications. Therefore, comprehending the nature and extent of return and volatility linkages across different financial markets can offer insights into diversification and hedging strategies for investors, as well as appropriate policy actions for regulatory bodies and governments. In terms of risk management, it is crucial to comprehend any linkages and transmission of risks between stock markets.

Many studies have demonstrated that increasing bilateral trade relations between countries can have a substantial impact on their stock market linkages. For example, Paramati et al. (2016) examined the impact of bilateral trade relations on the time-varying correlations of the stock markets of Australia and 10 Asian countries. Their results showed that increasing bilateral trade relations between Australia and Asian countries contributed to their stock market interdependence over time.

The economic relationship between Australia and Japan is underpinned by complementary strengths and needs. According to the Australian Bureau of Statistics, Japan was Australia's third-largest trading partner and second-largest export destination in 2021. Australia is the single largest supplier of both industrial raw materials and energy to Japan (Australian Bureau of Statistics 2021b, Tables 14a and 14b). Meanwhile, Japan ranks second only to the United States in terms of the stock of foreign direct investment (FDI) into Australia (Australian Bureau of Statistics 2021a, Table 2).

Australia and Japan have some common and distinct features in terms of their economic structures. The Australian economy is dominated by its service sector and is rich in natural resources, while Japan is one of the largest and most developed economies in the

Citation: Marumo, Kohei, and Steven Li. 2024. Downside Risk in Australian and Japanese Stock Markets: Evidence Based on the Expectile Regression. *Journal of Risk and Financial Management* 17: 189. <https://doi.org/10.3390/jrfm17050189>

Academic Editor: Thanasis Stengos

Received: 18 March 2024

Revised: 24 April 2024

Accepted: 28 April 2024

Published: 2 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

world. Japan has a well-educated, industrious workforce, and its large, affluent population makes it one of the world's biggest consumer markets. Manufacturing has been the most remarkable and internationally renowned feature of Japan's economic growth.

Both Australia and Japan play critical roles in the world financial markets. Australian Securities Exchange (ASX) is one of the world's leading financial market exchanges, offering a full suite of services, including listings, trading, clearing and settlement, across a comprehensive range of asset classes. It was created by the merger of the Australian Stock Exchange and the Sydney Futures Exchange in July 2006 and is one of the world's top-10 listed exchange groups measured by market capitalization¹.

The Tokyo Stock Exchange (TSE) is a stock exchange located in Tokyo, Japan. It is the fourth-largest stock exchange in the world by the aggregate market capitalization of its listed companies and largest in Asia. Tokyo is one of the world's foremost centers of finance.

The economic relationship between Australia and Japan is strong, and it is essential to have a comprehensive understanding of the transmission of risk between their financial markets. This study aims to contribute to the literature by examining the influence of downside risk between the two stock markets, using the Expectile Value-at-Risk (EVaR) as the risk measure. Specifically, the paper will apply an expectile regression approach to evaluate the EVaR of the stock market indices of both countries.

The expectile regression model that includes lagged returns and some common risk factors will be used to calculate the EVaR for each stock market. The model will also be used to assess the influence between the two markets in terms of downside risk. By doing so, the study will shed light on the interdependence of the two markets and how they respond to market shocks.

Overall, this research will provide valuable insights into the dynamics of risk transmission between the financial markets of Australia and Japan, which can help investors and policymakers better understand the potential risks and opportunities in these markets.

The remainder of the paper is organized as follows. Section 2 provides a review of relevant literature and emphasizes the significance of the study. Section 3 describes the research methodology used, along with the data and sample statistics. The empirical results and analysis are presented in Section 4, and concluding remarks are provided in Section 5.

2. Related Literature

In this section, we first describe the downside risk measures, then provide a review of the literature related to risk transmission between financial markets, with a particular focus on studies that employ the EVaR approach. This section will also highlight the importance of this study in advancing our understanding of the interdependence of the Australian and Japanese financial markets.

2.1. Measures of Risk

Risk is defined as the potential for an undesirable outcome (i.e., loss of value) resulting from an action or inaction. In finance, risk is the uncertainty surrounding the future value of an asset or portfolio of financial instruments.

Measuring and managing risk is crucial in the financial industry to ensure the stability and sustainability of the financial system. A good risk measure should accurately capture the potential for losses, but not be overly conservative, as this could restrict economic growth and investment opportunities. A variety of risk measures are used in finance, including Value-at-Risk (VaR), Expected Shortfall (ES), and the Expectile Value-at-Risk (EVaR), each with its own strengths and weaknesses.

The financial disasters of the early 1990s led to significant changes in the financial landscape, with Value-at-Risk (VaR) becoming a standard risk management tool in financial and insurance institutions since the 1990s. VaR measures the potential amount a portfolio can lose for a given probability level, defined as a tail quantile of the distribution of financial returns. However, VaR has some significant drawbacks, including ignoring most

of the risk's probability distribution and not being a sub-additive risk measure (Du and Escanciano 2016; Gordy and Juneja 2010), which can result in the measure for a portfolio being less than the sum of its components' measures.

Expected Shortfall (ES) has been proposed as a risk measure for regulatory frameworks in response to the limitations of Value at Risk (VaR) (Du and Escanciano 2016; Embrechts et al. 2014). Unlike VaR, which only considers the probability of losses exceeding a certain threshold, ES calculates the average of the VaR values below a specified level, thus including very unlikely losses with a significant impact on the portfolio value. ES is a coherent risk measure and has the property of subadditivity, thus rewarding diversification (Hult et al. 2012). In fact, Basel III² has recommended ES as a replacement for VaR in the new regulatory framework and raised the capital ratios required for financial institutions.

In addition to their use as risk measures, VaR and ES estimation can serve as the basis for portfolio optimization (Huang et al. 2010; Lwin et al. 2017). However, it is worth noting that ES is not elicitable, meaning that there is no scoring function for which the expectation is minimized by the true ES (Gneiting 2011). Nonetheless, the benefits of ES as a coherent risk measure have led to its increased use in regulatory frameworks and portfolio optimization.

The estimation and evaluation of Expected Shortfall (ES) has been challenging in practice, as noted by Rockafellar et al. (2014). However, Fissler et al. (2016) showed that ES and VaR are jointly elicitable. They also proposed a family of scoring functions that can be used to estimate and evaluate both VaR and ES simultaneously. This has enabled researchers such as Patton et al. (2019) and Taylor (2019) to estimate dynamic models for VaR and ES based on joint scoring functions. Additionally, Meng and Taylor (2020) introduced a novel approach to estimating these two risk measures using joint scoring functions based on intraday data. These methods have advanced our ability to estimate and evaluate VaR and ES accurately, despite the challenges involved in their practical implementation.

One approach to enhancing risk measures is by identifying and accounting for structural breaks in the markets. A prevalent method involves introducing dummy variables (Anjum and Malik 2020; Patra 2021). Nevertheless, for simplicity, we employ the conventional model in this study and defer addressing this issue to future work.

2.2. Risk Transmission between Financial Markets

As global financial markets are increasingly deregulated and integrated, the risk in one financial market or asset is very likely to be transmitted to other markets or assets. Understanding risk transmission between financial markets is important for asset allocation and risk management. It also has crucial implication for policy making that aims at reducing financial contagion and the instability it causes to the economy.

Risk transmission between financial markets has been the subject of numerous studies over the past few decades. The interconnectedness of global financial markets has become increasingly apparent since the 2008 Global Financial Crisis, where the spillover effects of one market's instability led to economic turmoil in other regions. A number of studies have sought to understand the mechanisms and channels through which risk is transmitted between financial markets.

Caporals et al. (2006) used a Vector Autoregression (VAR) model to examine the cross-market volatility spillovers between the US and European equity markets. Their results showed that spillovers were significant and that US markets were more influential in transmitting shocks to European markets than vice versa. This was corroborated by other studies, such as Forbes and Rigobon (2002) who found that US equity market volatility was a significant predictor of volatility in other major equity markets.

Later on, many studies investigate volatility spillover using multivariate GARCH models. For instance, Baele (2005) analyzed the volatility spillover effects in European equity markets. Li and Giles (2015) studied the volatility spillover effects between developed stock markets and Asian emerging stock markets. Vo and Tran (2020) investigated volatility spillovers from the US equity market to ASEAN stock markets. Zhong and Liu (2021) pro-

vided evidence on volatility spillovers between China and South-east Asian stock markets. These are just a few examples of studies that have examined the phenomenon of volatility spillovers across different equity markets.

However, in many applications, losses and gains are treated symmetrically in the measure of correlation and volatility spillover. In reality, investors have downside financial constraints and are downside risk averse, meaning they are more concerned with downside losses than upside gains. Therefore, these applications are inadequate in describing extreme bad scenarios when a significant decline in stock returns in one market generates panic and spreads to other markets, as seen in events such as the 1997 Asian financial crisis and the Global Financial Crisis in 2008.

To address these issues, some studies have distinguished extreme negative shocks from normal observations. For example, Hartmann et al. (2004) looked at extreme dependence between markets in distress periods, while Asgharian and Bengtsson (2006) and Asgharian and Nossman (2011) analyzed risk spillover with Poisson jumps. Hussain and Li (2018) considered the dependence structure between Chinese and other major stock markets based on the extreme value theory and copulas. A few studies (e.g., Adrian and Brunnermeier 2009; Bae et al. 2003; Christiansen and Rinaldo 2009) study extreme coexceedance or simultaneous extreme events of financial markets and financial institutions.

Further, a few studies have focused on risk transmission between stock markets using downside risk measures such as Expected Shortfall (ES) and Expectile Value-at-Risk (EVaR). Hong et al. (2009) provided a theoretical background for the binary response of extreme value spillover. Inspired by this concept of Granger causality in risk, Liu (2014) proposed a binary response model approach to measure and forecast extreme downside risk in Asia-Pacific markets, given information on extreme downside risks in the US and Japanese markets. In both studies, downside risk is measured by VaR. However, there is a need for more research on risk transmission between stock markets using downside risk measures such as ES and EVaR.

2.3. The Expectile Approach

The term “expectile” was first introduced by Newey and Powell (1987). One major advantage of using expectiles is that asymmetric least squares (ALS) regression is more computationally efficient than quantile regression. Computing the expectile is simpler than computing the quantile because the check-loss function is continuously differentiable. This has led to an increase in popularity in the use of expectiles in modelling risk measurement and transmission in recent years.

Expectiles provide a different approach to measuring risk as they rely on both tails of the distribution, unlike other measures such as expected shortfall and value-at-risk. Expected shortfall is sensitive to the magnitude of the lower tail returns, but only those of the lower tail. As such, it tends to be more conservative than other risk measures. VaR is not influenced by the magnitude of lower tail returns at all, increasing the chance that it underestimates the true risk involved. The expectile relies on more comprehensive information to measure risk.

The relationship between expectiles and quantiles has also encouraged the use of expectiles. Efron (1991) suggested quantiles can be estimated using the expectile. The one-to-one mapping between quantiles and expectiles has been empirically supported by Jones (1994), Abdous and Remillard (1995), and Yao and Tong (1996). Taylor (2008) estimated nonparametric value-at-risk and expected shortfall using expectiles. The emphasis has generally been placed on the computational efficiency of the expectile as compared to quantile regression, rather than on the virtues of the expectile as a risk measure. De Rossi and Harvey (2009) were the first to directly use the expectile by developing a spline-based nonparametric computation method for the time-varying expectile. Its use was still limited by the difficulty in interpreting the expectile and its asymmetry parameter. Direct use of the expectile as a risk measure has been discouraged by the expectiles’ lack of

interpretation. Kuan et al. (2009) addressed the issue of interpretability by giving a more intuitive definition for the expectiles in a financial risk setting.

Additionally, Zhang and Ma (2019) examine the connections between crude oil futures markets and stock markets through expectile regressions, demonstrating the effectiveness of this approach. In a separate study, Yao et al. (2021) employ expectile regressions to assess the risk of the Chinese FinTech index, exploring associated macroeconomic risk factors.

As our objective is to examine the downside risk connection among specific markets, we utilize the a priori specified model, detailed below. Notably, Xu et al. (2016) present an intriguing application of a nonparametric approach for model specification. Furthermore, Liao et al. (2019) introduce penalized expectile regressions; however, we opt for the conventional method for simplicity.

In sum, the expectile, which is the solution to asymmetric least squares regression, resolves many of the issues of previous risk measures. In addition to addressing the shortcomings of other risk measures such as Value-at-Risk and Expected Shortfall, expectiles have unique characteristics that make them very efficient and informative for measuring risk.

3. Methodology and Data

In this section, we will detail the research methodology used in the study, including the expectile regression model, EVaR calculations, and the common risk factors used in the model. This section will also describe the data sources and sample statistics used in the study.

3.1. Traditional Downside Risk Measures

A VaR with a confidence level of $1 - \alpha$, $\alpha \in (0, 1)$, represents the maximum potential loss for a given holding period with a probability of $1 - \alpha$, see e.g., Jorion (2000). VaR is essentially the negative of the α -th quantile of the return distribution and can be obtained by minimizing asymmetrically weighted mean absolute deviations. In this weighting scheme, positive deviations are assigned a weight of α , while negative deviations are assigned a weight of $1 - \alpha$. Bassett et al. (2004) argued that this asymmetric weighting scheme aligns with certain distorted probability assessments used in Choquet expected theory, which can describe pessimism.

To be specific, let Y denote an asset return with the distribution function F_Y . Given $\alpha \in (0, 1)$, the quantile based VaR (henceforth QVaR) is the negative of the α -th quantile of F_Y : $\text{QVaR}(\alpha) = -q(\alpha)$. It is well known that the α -th quantile can be obtained by minimizing asymmetrically weighted mean absolute deviations:

$$E[|\alpha - \mathbf{1}_{\{Y \leq q\}}| |Y - q|] \quad (1)$$

where $\mathbf{1}_{\{Y \leq q\}}$ is the indicator of the event $Y \leq q$. Thus, a QVaR is a natural product of an optimization problem with an asymmetric linear loss function. The first order condition of minimizing Equation (1) is $(1 - \alpha) \int_{-\infty}^q dF_Y(y) - \alpha \int_q^{\infty} dF_Y(y) = 0$, which implies

$$\alpha = \frac{\int_{-\infty}^q dF_Y(y)}{\int_{-\infty}^{\infty} dF_Y(y)} = \int_{-\infty}^q dF_Y(y). \quad (2)$$

This highlights that VaR has limitations as it depends solely on the frequency of more extreme losses without considering their magnitudes, unlike the lower partial moment. This may result in VaR being less sensitive to tail risk, meaning that two portfolios with the same VaR could have very different tail shapes. Additionally, VaR does not satisfy sub-additivity, so diversification may not decrease the risk measured by VaR, which is undesirable.

In contrast, expected shortfall (ES) considers the average of all potential losses beyond the VaR at a given confidence level, taking into account the magnitude of the loss.

However, ES only accounts for the conditional downside mean, and may therefore be overly conservative.

For an asset return Y with the distribution function F_Y and given $\alpha \in (0, 1)$, the ES of Y at level α is the mean of the α -tail distribution of Y (Rockafellar and Uryasev 2002; Tasche 2002), i.e.,

$$ES_\alpha(Y) = \int_{-\infty}^{\infty} y dF_{\alpha,Y}(y),$$

where $F_{\alpha,Y}$ is the α -tail distribution defined as (Rockafellar and Uryasev 2002)

$$F_{\alpha,Y}(y) = \begin{cases} 0, & \text{for } y < QVaR(\alpha) \\ \frac{F_Y(y) - \alpha}{1 - \alpha}, & \text{for } QVaR(\alpha) \leq y. \end{cases}$$

For a more detailed discussion, we refer to Kou and Peng (2016).

3.2. Expectile Based VaR

While the previously mentioned risk measures are still widely used, they have been criticized for their limitations. The expectile approach has emerged as a solution to many of these issues. Expectiles have unique properties that make them highly efficient and informative in measuring risk.

As opposed to quantiles, which only provide information on a specific percentile, expectiles capture information on a specific tail and are more sensitive to tail risks. Additionally, expectiles are robust to outliers and can be easily adapted to incorporate various risk factors. As such, the expectile regression model has become an increasingly popular tool in risk management and financial analysis. The EVaR of Y , an asset return with the expectile level θ , ($\theta \in (0, 1)$) is the negative of θ -th expectile of F_Y , the distribution function, which can be obtained as the negative of ν that minimizes the asymmetrically weighted quadratic loss function (Newey and Powell 1987),

$$E\left[|\theta - \mathbf{1}_{\{Y \leq \nu\}}|(Y - \nu)^2\right], \quad (3)$$

where $\mathbf{1}_{\{Y \leq \nu\}}$ is the indicator of the event $Y \leq \nu$. Let us denote such value of ν by $\nu(\theta)$. Then we have: $EVaR(\theta) = -\nu(\theta)$ with $0 < \theta < 1/2$.

The first-order condition of minimizing (3) is:

$$-(1 - \theta) \int_{-\infty}^{\nu} (y - \nu) dF_Y(y) - \theta \int_{\nu}^{\infty} (y - \nu) dF_Y(y) = 0 \quad (4)$$

which implies

$$\theta = \frac{\int_{-\infty}^{\nu} (y - \nu) dF_Y(y)}{\int_{-\infty}^{\nu} (y - \nu) dF_Y(y) - \int_{\nu}^{\infty} (y - \nu) dF_Y(y)} = \frac{\int_{-\infty}^{\nu} |y - \nu| dF_Y(y)}{\int_{-\infty}^{\infty} |y - \nu| dF_Y(y)}. \quad (5)$$

Hence, θ is the ratio of the deviations of Y below ν to the overall deviations of Y from ν , both weighted by the distribution function. Thus, $\nu(\theta)$ depends on both the extreme values and their probabilities (Kuan et al. 2009).

Lower values of θ indicate more risk aversion. Kuan et al. (2009) suggested that the EVaR can be thought of as a flexible VaR, since the tail probability associated with an expectile is not some static θ chosen at the outset, but changes with the underlying distribution.

Compared to the popular QVaR, EVaR is more sensitive to the extreme values of the distribution (Kuan et al. 2009; Xie et al. 2014). Furthermore, $\nu(\theta)$ can be used to calculate

ES according to the one-to-one mapping from α to θ . Yao and Tong (1996) show that $\theta(\alpha)$ and $q_\alpha(Y)$ satisfy

$$\theta(\alpha) = \frac{\alpha q_\alpha(Y) - \int_{-\infty}^{q_\alpha(Y)} y dF_Y(y)}{2E[Y] - 2 \int_{-\infty}^{q_\alpha(Y)} y dF_Y(y) - (1 - 2\alpha)q_\alpha(Y)}. \quad (6)$$

Hence, for any $\alpha \in (0, 1)$, $\theta(\alpha)$ is transformed by Equation (6), such that $v(\theta(\alpha))(Y) = q_\alpha(Y)$. ES can be obtained according to Taylor (2008) by,

$$ES_\alpha(Y) = -\left(1 + \frac{\theta}{(1 - 2\theta)\alpha}\right)v(\theta) + \frac{\theta}{(1 - 2\theta)\alpha}E[Y]. \quad (7)$$

3.3. Expectile Regression

Given a collection of k variables $X = (1, X_2, X_3, \dots, X_k)'$, in the information set $\mathcal{F}$, let $v(\theta|\mathcal{F})$ denote the θ -th expectile of Y conditional on X . We consider the linear specification $v(\theta|\mathcal{F}) = X'\beta(\theta)$ with $\beta(\theta)$ as a $k \times 1$ parameter vector.

When the data (y_t, x_t') , $x_t = (1, x_t^{(2)}, \dots, x_t^{(k)})'$ are available for $t = 1, \dots, T$, the linear specification can be expressed as:

$$v(\theta|\mathcal{F}_t) = x_t'\beta(\theta).$$

where $\mathcal{F}_t$ denotes the information set up to time t .

The parameter $\beta(\theta)$ can be estimated by the value of $\beta(\theta)$ that minimizes

$$\frac{1}{T} \sum_{t=1}^T |\theta - \mathbf{1}_{\{y_t - x_t'\beta(\theta) \leq 0\}}| (y_t - x_t'\beta(\theta))^2,$$

whose first order condition is

$$-\frac{1}{T} \sum_{t=1}^T |\theta - \mathbf{1}_{\{y_t - x_t'\beta(\theta) \leq 0\}}| x_t' (y_t - x_t'\beta(\theta)) = 0.$$

We can obtain such value by the usual iteration algorithm:

1. Let $\beta^{[0]}$ be an appropriate initial guess of the parameter $\beta(\theta)$.
2. Given $\beta^{[k]}$, the k th guess, $\beta^{[k+1]}$ is calculated as

$$\beta^{[k+1]} = \left(w^{[k]} \mathbf{x} \mathbf{x}'\right)^{-1} w^{[k]} \mathbf{x} y,$$

where

$$w^{[k]} = \begin{pmatrix} |\theta - \mathbf{1}_{\{y_1 - x_1'\beta^{[k]} \leq 0\}}| \\ \vdots \\ |\theta - \mathbf{1}_{\{y_T - x_T'\beta^{[k]} \leq 0\}}| \end{pmatrix}, \quad \mathbf{x} = (x_1, \dots, x_T), \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix}.$$

3. Repeat Step 2, for instance, until $|\beta^{[k+1]} - \beta^{[k]}|$ is small enough.

Let us denote by $\hat{\beta}(\theta)$ the estimate of the parameter $\beta(\theta)$ obtained by the algorithm above. Kuan et al. (2009) suggest that for large T the distribution of $\hat{\beta}(\theta)$ be approximated by the normal distribution with mean $\beta(\theta)$ and variance $T\hat{\Xi}(\theta)^{-1}\hat{V}(\theta)\hat{\Xi}(\theta)^{-1}$, where

$$\hat{\Xi}(\theta) = \frac{1}{T} \sum_{t=1}^T |\theta - \mathbf{1}_{\{y_t - x_t'\hat{\beta}(\theta) \leq 0\}}| x_t x_t', \quad \hat{V}(\theta) = \frac{1}{T} \sum_{t=1}^T |\theta - \mathbf{1}_{\{y_t - x_t'\hat{\beta}(\theta) \leq 0\}}|^2 (y_t - x_t'\hat{\beta}(\theta))^2 x_t x_t'.$$

We note that Kuan et al. (2009) provide some simple numerical examples, and the proof can be attributed to Newey and Powell (1987). Furthermore, expectile regression has seen active application in recent years within financial empirical research.

3.4. Variables

To investigate if the Japanese stock market is a risk factor that influences the Australian stock market, similar to Yao et al. (2021), we use the following variables:

- y = ASX 200 index return;
- $x^{(2)}$ = Nikkei 225 index return;
- $x^{(3)}$ = S&P 500 index return; and
- $x^{(4)}$ = CSI 300 index return.

We also use the specification

$$v(\theta|\mathcal{F}_t) = \left(1\ y_{t-1}^+ \ y_{t-1}^- \ y_{t-2}^+ \ y_{t-2}^- \ x_{t-1}^{(2)+} \ x_{t-1}^{(2)-} \ x_{t-1}^{(3)+} \ x_{t-1}^{(3)-} \ x_{t-1}^{(4)+} \ x_{t-1}^{(4)-}\right)\beta(\theta),$$

where we use the notations $u^+ = \max(u, 0)$ and $u^- = \min(u, 0)$.

To investigate if the Australian stock market is a risk factor that influences the Japanese stock market, we use the following variables:

- y = Nikkei 225 index return;
- $x^{(2)}$ = ASX 200 index return;
- $x^{(3)}$ = S&P 500 index return; and
- $x^{(4)}$ = CSI 300 index return.

In both cases, the CSI 300 and S&P 500 are considered as common risk factors to both the Australian and Japanese stock markets.

3.5. Data and Sample Statistics

The two largest economies in the world are the U.S. and China, and both countries have important economic relationships with Australia and Japan. As the U.S. stock market and the Chinese stock market are major players in the global market, they are expected to have significant influence on many other stock markets, including those of Australia and Japan. Therefore, in order to examine the risk transmission between the Australian and Japanese stock markets, it is reasonable to use the CSI 300 and S&P500 as the common risk factors in the expectile regression models. By using these two factors, we can also assess the spillover effects of the U.S. and Chinese markets on the Australian and Japanese markets, and better understand the risk transmission mechanism between them.

In this paper, we use the daily closing stock index data for Australia (ASX 200), Japan (Nikkei 225), U.S. (S&P 500), and China (CSI 300) over the past 10 years, from the beginning of October 2012 to end of September 2023. The exact dates vary depending on the operating days of each market. Table 1 presents the summary statistics for the four stock index returns over the sample period.

Table 1. Summary statistics of daily returns of the stock market indices.

Index	ASX 200	Nikkei 225	S&P 500	CSI 300
Mean ($\times 10^{-4}$)	1.18	3.24	3.73	1.79
Median ($\times 10^{-4}$)	6.5	7.81	5.74	2.23
Max ($\times 10^{-2}$)	6.77	7.73	8.97	6.5
Min ($\times 10^{-2}$)	−10.2	−8.25	−12.77	−9.15
S.dev ($\times 10^{-2}$)	0.97	1.26	1.11	1.4
Skew.	−1.05	−0.15	−0.81	−0.8
Kurt.	12.42	4.28	16.19	6.03
ADF	−1.92	−1.14	−0.81	−2.31
(<i>p</i> -value)	(0.32)	(0.70)	(0.82)	(0.17)

Based on Table 1, it can be observed that the CSI 300 index has the highest standard deviation, while the ASX 200 index has the lowest. This suggests that the Australian stock market was comparatively less volatile than the other three stock markets during the sample period. Additionally, the negative skewness and large kurtosis of the return distributions for all four stock indices indicate that they have a fat tail in the left part, with more extreme information. This characteristic makes the expectile regression model, which is more sensitive to the left tail, a suitable tool to measure tail risk.

All Augmented Dickey–Fuller (ADF) test p -values for the return distributions of the four stock indices are greater than 0.1. This suggests that the return series are stationary and that the expectile regression model can be effectively applied.

For illustrative purposes, Figure 1 displays the stock indices during the sample period from 1 October 2013, to 30 September 2023, with values normalized such that the value on 1 October 2013, is set to 100. It is evident that all indexes experienced significant fluctuations during the COVID-19 crisis.

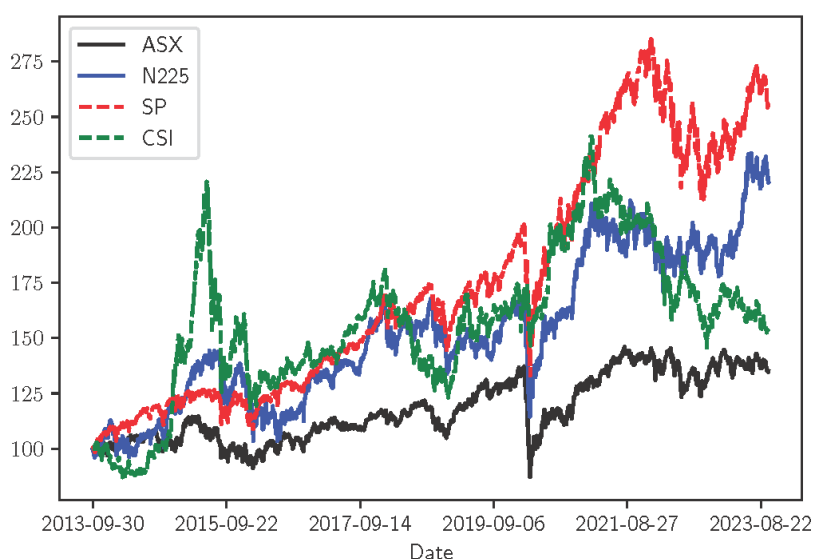


Figure 1. The evolution of the stock indices during the sample period.

4. Empirical Results

In this section, we present the empirical results and analysis of our study. We begin by examining the Expectile based Value at Risk (EVaR) for each stock market and then move on to assessing the influence of downside risk between the Australian and Japanese markets. We also analyze the impact of lagged returns and common risk factors on the EVaR and Expected Shortfall (ES) estimates.

4.1. The Effect of Risk Factors

For the expectile regression, the corresponding θ is calculated according to Equation (6) for a given α (1% or 5% in this paper). The final expectile regression results for the ASX 200 and Nikkei 225 indexes are reported in Table 2.

Table 2. The expectile regression results for ASX 200 and Nikkei 225 indices. “s.e.” denotes the standard error, ***, **, * label significance at 1%, 5%, and 10% levels, respectively.

Indices	ASX 200		Nikkei 225	
α	1%	5%	1%	5%
θ	0.67%	2.54%	0.3%	2.13%
Variables	Estimate (s.e.)	Estimate (s.e.)	Estimate (s.e.)	Estimate (s.e.)
Const.	−0.0081 (0.0014) ***	−0.0059 (0.0011) ***	−0.0206 (0.0022) ***	−0.0098 (0.0014) ***
y_{t-1}^+	−0.0781 (0.2363)	−0.0275 (0.1210)	−0.1201 (0.1240)	−0.0691 (0.0914)
y_{t-1}^-	0.3097 (0.2185)	0.1114 (0.2031)	−0.0314 (0.1513)	0.0498 (0.0820)
y_{t-2}^+	−0.7581 (0.1854) ***	−0.5408 (0.1566) ***	−0.2603 (0.1339) *	−0.2693 (0.1044) ***
y_{t-2}^-	0.6885 (0.2962) **	0.434 (0.1884) **	0.4659 (0.0887) ***	0.5004 (0.0849) ***
$x_{t-1}^{(2)+}$	−0.0758 (0.0884)	−0.0903 (0.0841)	0.0615 (0.3145)	−0.1439 (0.1737)
$x_{t-1}^{(2)-}$	0.0035 (0.0884)	−0.0166 (0.0781)	0.5495 (0.2405) **	0.4419 (0.1701) ***
S&P 500 $_{t-1}^+$	−0.2258 (0.1226)	−0.1205 (0.1708)	−0.1587 (0.9359)	0.3057 (0.1504) **
S&P 500 $_{t-1}^-$	0.7614 (0.1776) ***	0.6062 (0.1124) ***	0.6437 (0.2245) ***	0.5785 (0.1169) ***
CSI 200 $_{t-1}^+$	−0.1036 (0.1127)	−0.0871 (0.0593)	0.2205 (0.1588)	0.0381 (0.0759)
CSI 200 $_{t-1}^-$	0.0648 (0.0563)	0.0671 (0.0437)	−0.0404 (0.0499)	0.0297 (0.0498)

4.1.1. Results for ASX 200 Index

Regarding the expectile regression results for ASX 200, several observations can be made. First, the 0.01 expectile $\nu(\theta|\mathcal{F}_t)$ is significantly affected by y_{t-2}^+ and y_{t-2}^- with coefficients of -0.7581 and 0.6885, respectively. The 0.05 expectile $\nu(\theta|\mathcal{F}_t)$ is also affected by y_{t-2}^+ and y_{t-2}^- with the coefficients of -0.5408 and 0.4340, respectively. As for the tail risk $\text{EVar}(\theta) = -\nu(\theta|F_t)$, the results suggest that negative returns of the ASX 200 index two business days before tend to indicate a lower tail risk, which results in higher EVaR. Conversely, positive returns tend to suggest higher tail risk and lower EVaR. These findings imply that ASX 200 tends to correct overreactions in terms of downside tail risk.

Second, the effect of negative returns of the S&P 500 index previous day is significant on both the 0.01 and 0.05 expectile $\nu(\theta|\mathcal{F}_t)$ with the coefficients of 0.7614 and 0.6062, respectively, while the CSI 300 index is not.

Third, the effect of the Nikkei 225 index previous day is not significant. The results indicate that the transmission of risk from Japanese stock markets to Australian ones can be limited.

4.1.2. Results for Nikkei 225 Index

Regarding the Nikkei 225 index, Table 2 reveals a few interesting findings as well. First, both the 0.01 and 0.05 expectile $\nu(\theta|\mathcal{F}_t)$ is significantly affected by the lagged y_{t-2}^+ and y_{t-2}^- .

For the tail risk $\text{EVar}(\theta) = -\nu(\theta|\mathcal{F}_t)$, the results indicate that the negative past returns (y_{t-2}^-) of the Nikkei 225 index tends to suggest a lower tail risk and push the EVaR upward while the positive past returns (y_{t-2}^+) tends to suggest a higher tail risk and push the EVaR

downward. However, the negative past returns or positive past returns of lag 1 do not have any significant effect.

Second, the effect of the negative ASX 200 index previous day is significant, at 5% level on the 0.01 expectile $v(\theta|\mathcal{F}_t)$, and at 1% level on the 0.05 expectile $v(\theta|\mathcal{F}_t)$. The coefficients, 0.5495 for 0.01 expectile, and 0.4419 for 0.05 expectile, are positive. This implies that the negative return on ASX 200 previous day tends to result in reduction in EvaR. This contrasts to the findings in Section 4.1.1 and implies that the effect of ASX 200 on Nikkei 225 is stronger than the other way round in terms of downside tail risk.

Third, the negative return on the SP 500 index previous day is significant (at 1% level) on the 0.01 expectile $v(\theta|F_t)$ with the coefficient of 0.6437 and the 0.05 expectile $v(\theta|\mathcal{F}_t)$ with the coefficient of 0.5785. The CSI is not significant here.

4.1.3. The Comparison

Comparing the results of the ASX 200 and Nikkei 225, it can be concluded that the effect of Australian stock market on Japanese ones can be stronger than the other way round in terms of downside tail risk, reflecting the economic integration between Australia and Japan in the form of trade linkages and foreign direct investment. This finding is in line with the existing literature, such as Shamsuddin and Kim (2003).

It can be observed that the Nikkei 225 reacts less to past positive returns compared to the ASX 200. This result confirms the existence of a difference between the two stock markets in the mechanism of downside risk transmission, even though they are highly integrated.

4.2. EVaR Estimates

In order to test the accuracy of the model, the EVaR was estimated to analyze the past risk conditions of each stock index.

Figures 2 and 3 show the in-sample $v(\theta|F_t)$ for ASX 200 and Nikkei 225, respectively. Both figures display a line that covers mostly negative returns, and for both ASX 200 and Nikkei 225, the effectiveness of the model under 0.01 quantile is better than the 0.05 quantile level.

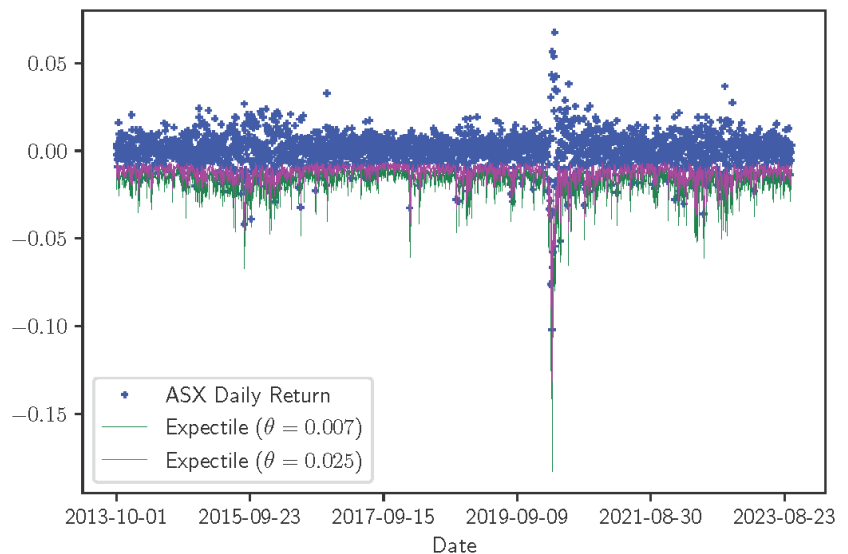


Figure 2. In-sample $v(\theta|\mathcal{F}_t)$ of ASX 200 index for $\theta = 0.0067$ ($\alpha = 0.01$) and 0.0254 ($\alpha = 0.05$). In-sample period: 1 October 2012 to 30 September 2023.

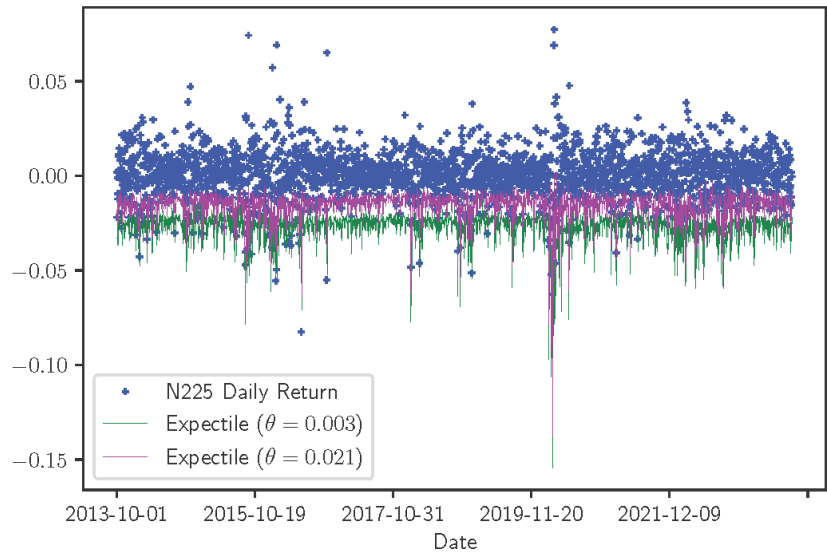


Figure 3. In-sample $\nu(\theta|\mathcal{F}_t)$ of Nikkei 225 index for $\theta = 0.0030$ ($\alpha = 0.01$) and 0.0213 ($\alpha = 0.05$). In-sample period: 1 October 2012 to 30 September 2023.

Interestingly, the EVaR of either index is larger during periods of high volatility in the return series, such as during the Chinese stock market crash in 2015 or at the beginning of the COVID-19 crisis in early 2020. However, the EVaR of Nikkei 225 is also high during a few other periods.

A few observations can be made from Table 3. The mean EVaR values of the ASX 200 index are 0.0186 and 0.0131 under 0.01 and 0.05 quantiles, respectively. These estimates suggest that the potential maximum loss of investing in the ASX 200 index one day ahead with 99% and 95% probabilities is 1.87% and 1.31%, respectively.

Table 3. EVaR and the tail probability of ASX 200 and Nikkei 225 indices.

α	1%			5%		
	Mean	Max	Tail Prob.	Mean	Max	Tail Prob.
ASX 200	0.0187	0.1829	2.729%	0.0131	0.1316	6.606%
Nikkei 225	0.0276	0.1544	1.432%	0.0163	0.1376	5.524%

The mean EVaR values of the Nikkei 225 index are 0.0276 and 0.0163 under 0.01 and 0.05 quantiles, respectively. These estimates suggest that the potential maximum loss of investing in the Nikkei 225 index one day ahead with 99% and 95% probabilities is 2.76% and 1.63%, respectively.

The mean EVaR values of the ASX 200 index under two quantiles are smaller than those of the Nikkei 225 index, while the maximum EVaR values of the ASX 200 index under 0.01 expectiles is larger than that of the Nikkei 225 index.

4.3. Out-of-Sample Estimation

To verify the robustness of the expectile regression results, we compare the in-sample and out-of-sample estimates. We use data from a seven-year in-sample period (1 October 2013, to 30 September 2020) to estimate the parameter $\hat{\beta}$ and obtain the expectiles $\nu(\theta|\mathcal{F}_t)$ for both in-sample and out-of-sample periods (1 October 2020, to 30 September 2023).

Table 4 compares the numbers of exceedances over the estimated expectile. It indicates that the ratio (or tail probability) remains largely consistent between the in-sample and

out-of-sample periods, except for the Nikkei 225 with $\alpha = 1\%$, where the out-of-sample estimate appears biased toward the conservative side.

Figures 4 and 5 display the in-sample and out-of-sample estimates of expectiles and realized returns.

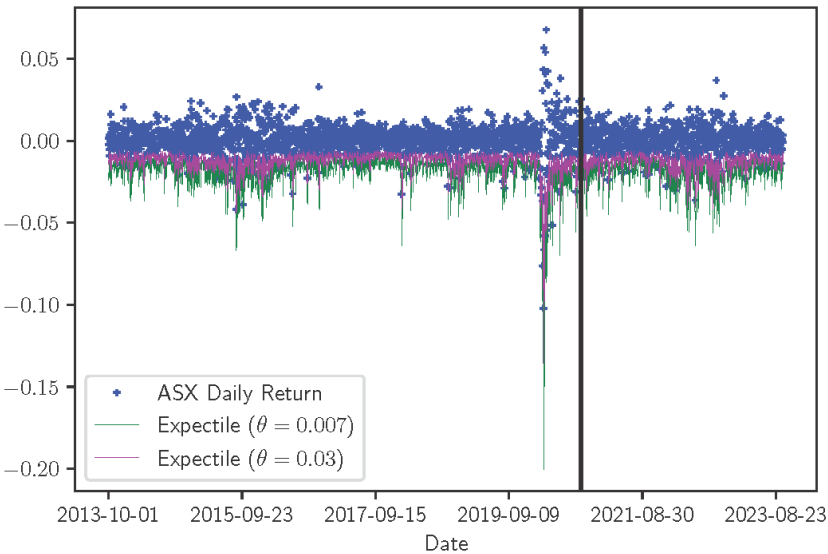


Figure 4. In-sample period: 1 October 2013 to 30 September 2020, the (left) side of the bar, and out-of-sample period: 1 October 2020 to 30 September 2023, the (right) side of the bar.

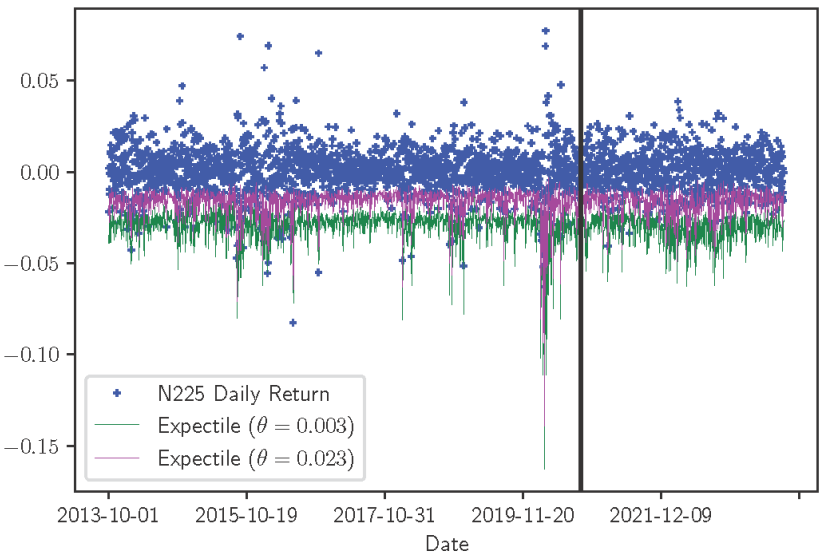


Figure 5. In-sample period: 1 October 2013 to 30 September 2020, the (left) side of the bar, and out-of-sample period: 1 October 2020 to 30 September 2023, the (right) side of the bar.

Table 4. Comparison of the number of exceedances over the estimated expectiles between in-sample and out-of -sample. In-sample period: 1 October 2013 to 30 September 2020. Out-of-sample period: 1 October 2020 to 30 September 2023.

Indices	ASX 200		Nikkei 225	
	α θ	1% 0.67%	5% 3.0%	1% 0.28% 5% 2.3%
In-sample		49/1770 (2.8%)	132/1770 (7.5%)	26/1709 (1.5%) 95/1709 (5.6%)
Out-of-sample		16/758 (2.1%)	45/758 (5.9%)	3/735 (0.4%) 29/735 (3.9%)

4.4. ES Estimates

Let us now consider the popular ES measure for evaluating the expected loss when investing in ASX 200 and Nikkei 225. Table 5 reports the mean and max ES values, as well as the tail probabilities under 1% and 5% probabilities.

The in-sample ES values for the ASX 200 index suggest that the mean risk when investing in it one day ahead is 3.12% and 2.01% with 1% and 5% probabilities, respectively. In contrast, the in-sample ES values for the Nikkei 225 index indicate that the mean risk is 3.58% and 2.34% with 1% and 5% probabilities, respectively. Thus, on average, the risk of investing in Nikkei 225 is higher than that of ASX 200. These results further confirm the suitability of the expectile model.

Additionally, the lower tail probability values in Table 5 indicate that our model is highly accurate.

Table 5. In-sample ES of ASX 200 and Nikkei 225 under 0.01 and 0.05 quantiles. See Figures 6 and 7 for the time plots.

α	1%			5%		
	Mean	Max	Tail Prob.	Mean	Max	Tail Prob.
ASX 200	0.0312	0.3061	0.237%	0.0201	0.2021	1.899%
Nikkei 225	0.0358	0.2005	0.245%	0.0234	0.1987	1.795%

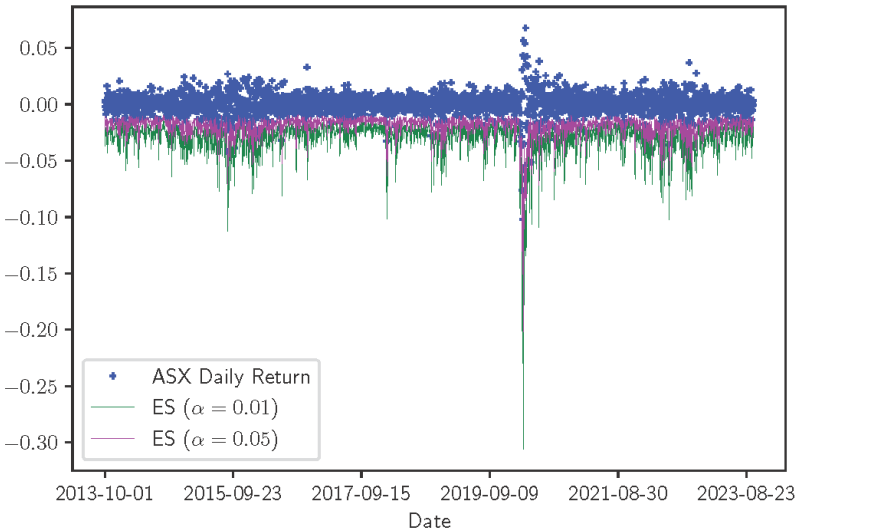


Figure 6. In-sample ES of ASX 200 under 0.01 and 0.05 quantiles.

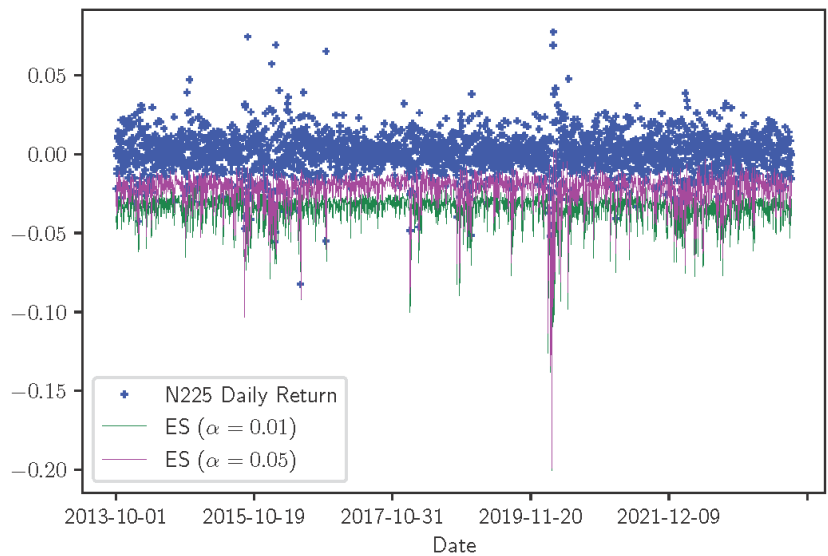


Figure 7. In-sample ES of Nikkei 225 under 0.01 and 0.05 quantiles.

5. Concluding Remarks

This study investigates the downside risk in Australian and Japanese stock markets and examines whether the two markets influence each other in terms of downside risk. To accomplish this, we use a novel approach based on the expectile regression model that accounts for both lagged returns and macroeconomic risk factors to calculate the EVaR of the ASX 200 index and Nikkei 225 index and evaluate their mutual influence.

Our analysis utilizes a ten-year sample (October 2013 to September 2023) of daily returns of the ASX 200 index and Nikkei 225 index. The EVaRs for both ASX 200 index and Nikkei 225 index are significantly influenced by negative lagged returns and returns from the S&P 500 index, while those of CSI 300 index is not significant. Moreover, the EVaR and ES of the ASX 200 index are on average smaller than those of the Nikkei 225 index.

Our findings suggest that both the Australian and Japanese stock markets are affected by their past development and international stock markets. Furthermore, our analysis indicates that the ASX 200 index has a significant impact on Nikkei 225 index downside tail risk, consistent with their close economic relationship. However, the ASX 200 index responds less to past negative or positive returns than the Nikkei 225 index. Our study's outcomes can have practical implications for investment and risk management in the two stock markets. In risk management practice, the measure of downside tail risk is often linked to the required capital level. Consequently, our findings can significantly enhance the effectiveness of the capital allocation policies.

Finally, it is worth noting that this research offers avenues for extension in several directions. First, we note that this research has a narrow focus, and it may be more meaningful to consider pairs of many other stock markets. Thus, we can broaden the scope by investigating other stock markets and commodity markets. Second, we can enhance our research methodology by, for example, identifying and accounting for structural breaks in the markets. Given the focus of this paper, we defer these aspects to future research.

Author Contributions: Conceptualization, S.L.; methodology, S.L. and K.M.; software, K.M.; validation, S.L. and K.M.; formal analysis, S.L. and K.M.; data curation, S.L. and K.M.; writing—original draft preparation, S.L. and K.M.; writing—review and editing, S.L. and K.M.; visualization, K.M.; funding acquisition, K.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by JSPS KAKENHI #20K01765.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Only publicly available market data were used in the study.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

¹ <https://www.asx.com.au/about>, accessed on 10 September 2023.

² <https://www.bis.org/bcbs/basel3.htm>, accessed on 10 September 2023.

References

- Abdous, Belkacem, and Bruno Remillard. 1995. Relating quantiles and expectiles under weighted-symmetry. *Annals of the Institute of Statistical Mathematics* 47: 371–84. [CrossRef]
- Adrian, Tobias, and Markus K. Brunnermeier. 2009. *CoVaR*. Technical Report 348. New York: Federal Reserve Bank of New York.
- Anjum, Hassan, and Farooq Malik. 2020. Forecasting risk in the us dollar exchange rate under volatility shifts. *The North American Journal of Economics and Finance* 54: 101257. [CrossRef]
- Asgharian, Hossein, and Christoffer Bengtsson. 2006. Jump spillover in international equity markets. *Journal of Financial Econometrics* 4: 167–203. [CrossRef]
- Asgharian, Hossein, and Marcus Nossman. 2011. Risk contagion among international stock markets. *Journal of International Money and Finance* 30: 22–38. [CrossRef]
- Australian Bureau of Statistics. 2021a. *International Investment Position, Australia: Supplementary Statistics*. Canberra: Australian Bureau of Statistics.
- Australian Bureau of Statistics. 2021b. *International Trade in Goods and Services, Australia*. Canberra: Australian Bureau of Statistics.
- Bae, Kee-Hong, G. Andrew Karolyi, and René M. Stulz. 2003. A new approach to measuring financial contagion. *Review of Financial Studies* 16: 717–63. [CrossRef]
- Baele, Lieven. 2005. Volatility spillover effects in european equity markets. *Journal of Financial and Quantitative Analysis* 40: 373–401. [CrossRef]
- Bassett, Gilbert W., Jr., Roger Koenker, and Gregory Kordas. 2004. Pessimistic portfolio allocation and choquet expected utility. *Journal of Financial Econometrics* 2: 477–92. [CrossRef]
- Caporale, Guglielmo Maria, Nikitas Pittis, and Nicola Spagnolo. 2006. Volatility transmission and financial crises. *Journal of Economics and Finance* 30: 376–90. [CrossRef]
- Christiansen, Charlotte, and Angelo Rinaldo. 2009. Extreme coexceedances in new eu member states' stock markets. *Journal of Banking and Finance* 33: 1048–57. [CrossRef]
- De Rossi, Giuliano, and Andrew Harvey. 2009. Quantiles, expectiles, and splines. *Journal of Econometrics* 152: 179–85. [CrossRef]
- Du, Zaichao, and Juan Carlos Escanciano. 2016. Backtesting expected shortfall: Accounting for tail risk. *Management Science* 63: 940–58. [CrossRef]
- Efron, Bradley. 1991. Regression percentiles using asymmetric squared error loss. *Statistica Sinica* 1: 93–125.
- Embrechts, Paul, Giovanni Puccetti, Ludger Rüschendorf, Ruodu Wang, and Antonela Beleraj. 2014. An academic response to basel 3.5. *Risks* 2: 25–48. [CrossRef]
- Fissler, Tobias, Johanna F. Ziegel, and Tilmann Gneiting. 2016. Expected shortfall is jointly elicitable with value at risk—Implications for backtesting. *Risk*, 58–61.
- Forbes, Kristin J., and Roberto Rigobon. 2002. No contagion, only interdependence: Measuring stock market comovements. *Journal of Finance* 57: 2223–61. [CrossRef]
- Gneiting, Tilmann. 2011. Making and evaluating point forecasts. *Journal of the American Statistical Association* 106: 746–62. [CrossRef]
- Gordy, Michael B., and Sandeep Juneja. 2010. Nested simulation in portfolio risk measurement. *Management Science* 56: 1833–48. [CrossRef]
- Hartmann, Philipp, Stefan Straetmans, and C. G. de Vries. 2004. Asset market linkages in crisis periods. *The Review of Economics and Statistics* 86: 313–26. [CrossRef]
- Hong, Yongmiao, Yanhui Liu, and Shouyang Wang. 2009. Granger causality in risk and detection of extreme risk spillover between financial markets. *Journal of Econometrics* 150: 271–87. [CrossRef]
- Huang, Dashan, Shushang Zhu, Frank J. Fabozzi, and Masao Fukushima. 2010. Portfolio selection under distributional uncertainty: A relative robust cvar approach. *European Journal of Operational Research* 203: 185–94. [CrossRef]
- Hult, Henrik, Filip Lindskog, Ola Hammarlid, and Carl-Johan Rehn. 2012. *Risk and Portfolio Analysis: Principles and Methods*. New York: Springer.
- Hussain, Saiful Izzuan, and Steven Li. 2018. The dependence structure between chinese and other major stock markets using extreme values and copulas. *International Review of Economics and Finance* 56: 421–37. [CrossRef]
- Jones, M. Chris. 1994. Expectiles and m-quantiles are quantiles. *Statistics Letters* 20: 149–53. [CrossRef]
- Jorion, Philippe. 2000. *Value at Risk: The New Benchmark for Managing Financial Risk*. Chicago: McGraw-Hill.

- Kou, Steven, and Xianhua Peng. 2016. On the measurement of economic tail risk. *Operations Research* 64: 1056–72. [CrossRef]
- Kuan, Chung-Ming, Jin-Huei Yeh, and Yu-Chin Hsu. 2009. Assessing value at risk with care, the conditional autoregressive expectile models. *Journal of Econometrics* 150: 261–70. [CrossRef]
- Liao, Lina, Cheolwoo Park, and Hosik Choi. 2019. Penalized expectile regression: An alternative to penalized quantile regression. *Annals of the Institute of Statistical Mathematics* 71: 409–38. [CrossRef]
- Liu, Lu. 2014. Extreme downside risk spillover from the united states and japan to asia-pacific stock markets. *International Review of Financial Analysis* 33: 39–48. [CrossRef]
- Li, Yanan, and David E. Giles. 2015. Modelling volatility spillover effects between developed stock markets and asian emerging stock markets. *International Journal of Finance and Economics* 20: 155–77. [CrossRef]
- Lwin, Khin T., Rong Qu, and Bart L. MacCarthy. 2017. Mean-var portfolio optimization: A nonparametric approach. *European Journal of Operational Research* 260: 751–66. [CrossRef]
- Meng, Xiaochun, and James W. Taylor. 2020. Estimating value-at-risk and expected shortfall using the intraday low and range data. *European Journal of Operational Research* 280: 191–202. [CrossRef]
- Newey, Whitney K., and James L. Powell. 1987. Asymmetric least squares estimation and testing. *Econometrica* 55: 819–47. [CrossRef]
- Paramati, Sudharshan Reddy, Rakesh Gupta, and Kishore Tandon. 2016. Dynamic analysis of time-varying correlations and cointegration relationship between australia and frontier equity markets. *International Journal of Business and Emerging Markets* 8: 121–45. [CrossRef]
- Patra, Saswat. 2021. Revisiting value-at-risk and expected shortfall in oil markets under structural breaks: The role of fat-tailed distributions. *Energy Economics* 101: 105452. [CrossRef]
- Patton, Andrew J., Johanna F. Ziegel, and Rui Chen. 2019. Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics* 211: 388–413. [CrossRef]
- Rockafellar, R. Tyrrell, and Stanislav Uryasev. 2002. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance* 26: 1443–71. [CrossRef]
- Rockafellar, R. Terry, Johannes O. Royset, and Sofia I. Miranda. 2014. Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional value-at-risk. *European Journal of Operational Research* 234: 140–54. [CrossRef]
- Shamsuddin, Abul F. M., and Jae H. Kim. 2003. Integration and interdependence of stock and foreign exchange markets: An australian perspective. *Journal of International Financial Markets, Institutions and Money* 13: 237–54. [CrossRef]
- Tasche, Dirk. 2002. Expected shortfall and beyond. *Journal of Banking and Finance* 26: 1519–33. [CrossRef]
- Taylor, James W. 2008. Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics* 6: 231–52. [CrossRef]
- Taylor, James W. 2019. Forecasting value at risk and expected shortfall using a semi-parametric approach based on the asymmetric laplace distribution. *Journal of Business and Economic Statistics* 37: 121–33. [CrossRef]
- Vo, Xuan Vinh, and Thi Tuan Anh Tran. 2020. Modelling volatility spillovers from the us equity market to asean stock markets. *Pacific-Basin Finance Journal* 59: 101246. [CrossRef]
- Xie, Shangyu, Yong Zhou, and Alan T. K. Wan. 2014. A varying coefficient expectile model for estimating value at risk. *Journal of Business and Economic Statistics* 32: 576–92. [CrossRef]
- Xu, Qifa, Xi Liu, Cuixia Jiang, and Keming Yu. 2016. Nonparametric conditional autoregressive expectile model via neural network with applications to estimating financial risk. *Applied Stochastic Models in Business and Industry* 32: 882–908. [CrossRef]
- Yao, Qiqli, and Howell Tong. 1996. Asymmetric least squares regression and estimation: A nonparametric approach. *Journal of Nonparametric Statistics* 6: 273–92. [CrossRef]
- Yao, Yinhong, Jianping Li, and Xiaolei Sun. 2021. Measuring the risk of chinese fintech industry: Evidence from the stock index. *Finance Research Letters* 39: 101564. [CrossRef]
- Zhang, Yue-Jun, and Shu-Jiao Ma. 2019. How to effectively estimate the time-varying risk spillover between crude oil and stock markets? evidence from the expectile perspective. *Energy Economics* 84: 104562. [CrossRef]
- Zhong, Yi, and Jiapeng Liu. 2021. Correlations and volatility spillovers between china and southeast asian stock markets. *The Quarterly Review of Economics and Finance* 81: 57–69. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

Analyzing Portfolio Optimization in Cryptocurrency Markets: A Comparative Study of Short-Term Investment Strategies Using Hourly Data Approach

Sonal Sahu ¹, José Hugo Ochoa Vázquez ¹, Alejandro Fonseca Ramírez ² and Jong-Min Kim ^{3,*}

¹ Department of Accounting and Finance, Campus Guadalajara, Tecnológico de Monterrey, Zapopan 45201, Mexico; sonalsahu@tec.mx (S.S.); hugoochoa@tec.mx (J.H.O.V.)

² EGADE Business School, Tecnológico de Monterrey, Monterrey 66269, Mexico; afonseca@tec.mx

³ Division of Science and Mathematics, University of Minnesota-Morris, Morris, MN 56267, USA

* Correspondence: jongmink@morris.umn.edu

Abstract: This paper investigates portfolio optimization methodologies and short-term investment strategies in the context of the cryptocurrency market, focusing on ten major cryptocurrencies from June 2020 to March 2024. Using hourly data, we apply the Kurtosis Minimization methodology, along with other optimization strategies, to construct and assess portfolios across various rebalancing frequencies. Our empirical analysis reveals significant volatility, skewness, and kurtosis in cryptocurrencies, highlighting the need for sophisticated portfolio management techniques. We discover that the Kurtosis Minimization methodology consistently outperforms other optimization strategies, especially in shorter-term investment horizons, delivering optimal returns to investors. Additionally, our findings emphasize the importance of dynamic portfolio management, stressing the necessity of regular rebalancing in the volatile cryptocurrency market. Overall, this study offers valuable insights into optimizing cryptocurrency portfolios, providing practical guidance for investors and portfolio managers navigating this rapidly evolving market landscape.

Keywords: portfolio optimization; Sharpe's; kurtosis minimization; rebalancing frequency; diversification; short-term investment strategy

Citation: Sahu, Sonal, José Hugo Ochoa Vázquez, Alejandro Fonseca Ramírez, and Jong-Min Kim. 2024. Analyzing Portfolio Optimization in Cryptocurrency Markets: A Comparative Study of Short-Term Investment Strategies Using Hourly Data Approach. *Journal of Risk and Financial Management* 17: 125. <https://doi.org/10.3390/jrfm17030125>

Academic Editor: Thanasis Stengos

Received: 29 February 2024

Revised: 16 March 2024

Accepted: 18 March 2024

Published: 20 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Cryptocurrencies are reshaping the financial landscape, poised to revolutionize the global financial system (Rajharia 2023). These digital assets garner attention from policy-makers, investors, and academics as a burgeoning alternative asset class (Jana et al. 2024), exhibiting unprecedented growth and establishing a significant international foothold (Subramoney et al. 2023). According to CoinMarketCap (2024), the current market capitalization of cryptocurrency is US \$2.72 trillion, showing a 3.67% increase from the previous day and highlighting the burgeoning interest in this domain (Krishna et al. 2023).

Cryptocurrency, characterized as a digital or virtual currency devoid of traditional backing like precious metals or commodities (Khan 2022), presents a unique challenge in understanding inter- and intra-asset dependencies among key financial variables (Zhang 2024). A comprehensive understanding of cryptocurrencies is imperative to expedite their broader adoption within the financial ecosystem (Shahzad 2024), particularly given their dynamic role within the evolving financial landscape (Wang 2024). Butterfill (2024) reported record weekly capital inflows totaling USD 2.7 billion in the first week of March 2024, bringing the total inflow year-to-date to USD 10.3 billion, nearing the record USD 10.6 billion inflow for the whole of 2021. This highlights the growing adoption and interest in cryptocurrencies.

The rapid expansion of cryptocurrencies is unfolding amidst an evolving regulatory framework, necessitating a nuanced approach to investment strategies. Regula-

tory considerations, encompassing consumer protection, financial stability, and taxation, loom large in the cryptocurrency space. Consequently, understanding interdependencies in cryptocurrency markets is indispensable for informed decision-making (Alipour and Charandabi 2023).

The genesis of cryptocurrencies traces back to Bitcoin's emergence in 2009, catalyzing a tumultuous evolution of the cryptocurrency market (ElBahrawy et al. 2017). Originally conceived to cater to the unbanked, reduce costs and energy consumption, and eliminate financial intermediaries (McMorrow and Esfahani 2021), cryptocurrencies have morphed into a distinct species within the financial instrument ecosystem. Characterized by decentralization, volatility, and high risk–reward potential, cryptocurrencies have ascended in global financial significance, attracting investors seeking diversification and short-term gains (Mungo et al. 2023).

Effective portfolio management emerges as pivotal in mitigating risks and maximizing returns in cryptocurrency investments. The dynamic nature of cryptocurrencies necessitates a thorough analysis to inform policymakers and regulatory bodies of their evolving role as investment assets. Cryptocurrencies offer potential avenues for risk diversification and portfolio optimization (Araujo et al. 2023), improving the risk–return trade-off and enhancing portfolio efficiency (Trimborn et al. 2019).

This study's novelty lies in its utilization of the top ten cryptocurrencies based on market capitalization. Our objective was to investigate the effectiveness of utilizing kurtosis in the portfolio construction framework, specifically by minimizing the kurtosis of the portfolio. This approach will be compared with the traditional Sharpe's maximization technique to determine which method performs better in the realm of cryptocurrency. Furthermore, we implemented five short-term strategies (half-week, one-week, two-week, four-week, and eight-week) for portfolio construction. Each strategy was tested using three different methodologies: naive portfolio construction, Sharpe maximization, and kurtosis minimization. This comprehensive approach allowed us to assess which short-term strategy aligns best with the volatile nature of cryptocurrencies.

To evaluate the performance of each portfolio, we considered both holding period returns and annualized returns. Statistical analysis techniques were then applied to the data to identify significant trends or patterns. This systematic analysis provided insights into the efficacy of different portfolio construction methods and short-term strategies within the cryptocurrency market.

The subsequent sections of this paper are structured as follows: in Section 2 is a literature review; Section 3 expounds on our empirical analysis; Section 4 presents the results and discussion; and, finally, Section 5 concludes our study.

2. Literature Review

The literature on cryptocurrency portfolio management strategies provided insights into various methodologies and their implications in cryptocurrency markets. Gupta et al. (2022) explored reducing portfolio risk using cryptocurrencies, highlighting the associated challenges. Silva et al. (2022) critically evaluated cryptocurrency trading algorithms, identifying areas for further exploration. Lorenzo and Arroyo (2023) compared risk-based portfolio allocation on crypto asset subsets, enriching the portfolio management literature. Symitsi and Chalvatzis (2019) quantified the economic gains achievable through cryptocurrency portfolio strategies. Previous research has explored interlinkages between major cryptocurrencies (Yousaf and Ali 2020), portfolio diversification in emerging markets (Letho et al. 2022), and the impact of volatility on cryptocurrency investment (Zaretta and Pangestuti 2023), offering insights into market dynamics.

The evolving regulatory landscape significantly influences investment strategies, impacting safety perceptions, risk characteristics, and market dynamics. Conlon et al. (2020) illustrated how regulatory developments during the COVID-19 pandemic affected the role of cryptocurrencies as safe havens for equity markets. Sukumaran et al. (2022) focused on regulatory developments in specific jurisdictions, while Waspada et al. (2022) stressed

the importance of regulation in managing risks associated with emerging crypto assets. Ethical considerations in cryptocurrency investments are crucial for responsible practices. Bagus and Horra (2021) undertook an ethical assessment of cryptocurrencies, highlighting considerations associated with investment decisions. Insights from the behavioral finance literature aid in understanding investors' psychological factors and decision-making in cryptocurrency investments (Ballis and Verousis 2022; Zhao and Zhang 2021; Choi and Feinberg 2022).

Transitioning to portfolio optimization methodologies, incorporating three distinct methodologies—naïve portfolio construction, Sharpe ratio maximization, and kurtosis minimization—offers a comprehensive approach that considers various aspects of risk and return (Hanif et al. 2023). Naïve portfolio construction involves equally weighting the assets in a portfolio and serves as a straightforward baseline strategy (Chen et al. 2022). Conversely, Sharpe ratio maximization aims to enhance the risk-adjusted return of a portfolio, effectively balancing risk and return (Dabbous et al. 2022). Lastly, kurtosis minimization seeks to mitigate the impact of extreme returns, bolstering portfolio resilience against market volatility (Susilo et al. 2020).

By integrating these methodologies, investors can obtain a holistic perspective on portfolio optimization. Naïve portfolio construction provides simplicity and serves as an initial step towards diversification, while Sharpe ratio maximization underscores the optimization of risk-adjusted returns, and kurtosis minimization concentrates on managing extreme risks. Together, these methodologies address various facets of portfolio performance, offering a well-rounded approach to constructing efficient and robust investment portfolios.

Short-term strategies and optimization methodologies play a crucial role in navigating the dynamics of the cryptocurrency market. Mean-variance optimization techniques provide valuable insights into managing short-term portfolios (Inci and Lagasse 2019), while robust risk management strategies help to mitigate short-term volatility (Kozlovskiy et al. 2022). The analysis of hourly data assists in understanding trading patterns, market efficiency, and the co-movements of volatility in cryptocurrency markets (Petukhina et al. 2020; Zhang et al. 2020).

Employing five short-term strategies across various durations—half-week, one-week, two-week, four-week, and eight-week—allows for a comprehensive approach to portfolio construction in cryptocurrencies. Each strategy serves a distinct purpose: the half-week strategy enables investors to rapidly adjust to capitalize on immediate market movements (Kong and Xu 2023), while the one-week strategy provides a slightly longer timeframe for evaluating trends (Gu 2022). Transitioning to a two-week strategy strikes a balance between agility and stability, capturing unfolding trends (Liu 2022). Extending to a four-week strategy enables investors to conduct a deeper analysis of performance and risk management (Zhang 2023). Finally, the eight-week strategy offers a longer-term perspective, facilitating strategic decision-making based on sustained trends (Ruban et al. 2022). By integrating these strategies, investors can tailor their approaches to different market conditions, optimizing risk-adjusted returns (Kong and Xu 2023).

The analysis of relationships and dependencies between cryptocurrencies through methodologies such as logarithmic returns and correlation matrices aids in portfolio diversification and risk management strategies (Scagliarini et al. 2022; Plerou et al. 2002). Testing strategies across different time horizons captures distinct aspects of their performance, emphasizing the importance of considering different time horizons in strategy testing (Nasir and Wahab 2022). Additionally, the buy-and-hold strategy exhibits variability in performance in cryptocurrency markets compared to other trading rules (Adrianus and Soekarno 2018).

The existing literature on cryptocurrency portfolio management strategies offers a comprehensive understanding of digital assets within the global financial system. Effective portfolio management strategies are crucial as cryptocurrencies disrupt traditional financial markets. The studies reviewed underscore the importance of considering interdependen-

cies among major cryptocurrencies, volatility's impact on investment decisions, and the significance of short-term strategies in navigating market dynamics.

Our study addresses a gap in the existing literature by examining the effectiveness of various portfolio optimization methodologies in short-term investments using a selection of ten cryptocurrencies. While previous studies have investigated portfolio management strategies and optimization techniques in cryptocurrency markets, we have not specifically focused on comparing different short-term investment strategies and their corresponding portfolio optimization methods. We aim to fill this gap by exploring which short-term investment strategy would be most suitable for investors in the cryptocurrency market and, within each strategy, identifying the optimal portfolio optimization approach. Through this analysis, we seek to provide valuable insights for investors looking to optimize their cryptocurrency portfolios for short-term gains.

3. Empirical Analysis

3.1. Data and Methodology

The hourly closing prices of the top ten cryptocurrencies by market capitalization—Binance (BNB), Bitcoin (BTC), Cardano (ADA), Chainlink (LINK), Dogecoin (DOGE), Ethereum (ETH), Litecoin (LTC), Polygon (MATIC), Ripple (XRP), and Tron (TRX)—were used to conduct this study. The data used for the study ranged from 1 June 2020 to 13 March 2024. All the data used for the analysis were downloaded from CoinMarketCap (<https://coinmarketcap.com/coins/>), (accessed on 11 March 2024). The reason for choosing this period was to incorporate the effects of the post-COVID-19 period.

Market capitalization serves as a vital metric for assessing the value and size of a cryptocurrency (Liu et al. 2022). Unlike individual cryptocurrency prices, which may not reflect the total value accurately, market capitalization provides a more comprehensive measure, allowing for comparisons with other cryptocurrencies. This concept mirrors the stock market, where market capitalization is calculated by multiplying the current stock price by the total number of shares outstanding. In the cryptocurrency realm, market capitalization is determined by multiplying the circulating supply (the number of coins available to the public) by the current price. Due to its significance, market capitalization is often a primary criterion for selecting cryptocurrencies. However, it should also be considered independently during portfolio optimization to ensure a well-rounded approach for practitioners.

Table 1 presents summary statistics for hourly cryptocurrency returns, revealing well-known financial stylized facts such as negative skewness and positive excess kurtosis. Out of the ten cryptocurrencies analyzed, eight showed negative average hourly returns, indicating a loss in value over the analyzed period. However, these average returns were close to 0%, with Polygon having the highest positive return and Litecoin showing the lowest negative return, as well as the largest range of 67.27%. Tron had the highest hourly median return at 0.019%.

The median and mean values for the sample were nearly identical, differing by less than 0.02%. The skewness, which measures the symmetry of the data distribution, varied among the cryptocurrencies, with all currencies skewed towards the lower tail and Cardano exhibiting the highest skewness in the sample.

The volatility, measured using the standard deviation and coefficient of variation, remained high for all cryptocurrencies, with the absolute values of the coefficient of variation ranging from 56.58 (Litecoin) to 493.69 (Binance). None of the currencies demonstrate normality at a 99% confidence level according to the Jarque–Bera test, as evidenced by consistently low p -values (<0.01). Additionally, all cryptocurrencies exhibited high kurtosis, with values ranging from 11.15 (Bitcoin) to 117.68 (Litecoin).

The combination of high volatility, skewness, and kurtosis in most currencies highlights the importance of a methodology that enables investors to understand currency behavior and construct improved portfolios by actively managing factors such as kurtosis.

Table 1. Shows the descriptive statistics of hourly returns of the ten cryptocurrencies.

Descriptive Statistics	Binance	Bitcoin	Cardano	Chainlink	Dogecoin	Ethereum	Litecoin	Polygon	Ripple	Tron
Minimum	−24.67%	−8.00%	−13.24%	−20.19%	−19.64%	−14.77%	−45.17%	−23.04%	−21.23%	−14.89%
Maximum	24.68%	9.34%	11.28%	11.37%	15.80%	7.37%	22.10%	19.33%	24.30%	9.27%
Range	49.35%	17.34%	24.52%	31.56%	35.44%	22.14%	67.27%	42.37%	45.53%	24.16%
Mean	0.00%	−0.01%	−0.01%	−0.02%	−0.02%	0.00%	−0.02%	0.01%	−0.02%	−0.01%
Median	0.01%	0.00%	−0.01%	0.00%	−0.01%	0.01%	0.01%	0.00%	0.02%	0.02%
Standard Deviation	1.32%	0.88%	1.34%	1.51%	1.56%	1.15%	1.38%	1.94%	1.48%	1.20%
Coefficient of Variation	493.69	−92.98	−126.87	−77.98	−69.19	−268.19	−56.58	230.49	−89.70	−86.95
Skewness	−0.26	−0.26	−0.15	−0.67	−0.53	−0.52	−3.68	0.40	−0.40	−0.60
Kurtosis	38.31	11.15	11.52	13.41	14.79	11.62	117.68	17.07	30.97	15.40
Jarque–Bera Test (<i>p</i> -value)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Source: Elaborated by the author.

We started by calculating the returns through the log difference of consecutive hourly closing prices of cryptocurrencies, following the method described by Akyildirim et al. (2021). This calculation followed the following equation:

$$R_n = (In\ CP_n - In\ CP_{(n-1)}) \times 100 \tag{1}$$

In this equation, R_n denotes returns at the n th hour as a percentage; CP_n denotes the closing price at the n th hour; $CP_{(n-1)}$ denotes the closing price on the previous trading hour; and $\ln$ is a natural log.

After analyzing the returns statistics, the next step is to create a simple portfolio. This involved allocating an equal fraction of $1/N$ of our total wealth to each of the N assets available for investment, following the ‘ $1/N$ ’ allocation rule. This approach is commonly recommended as a reference portfolio. In this study, we assigned 10% to each of the 10 assets in the portfolio. We conducted empirical investigations using two distinct portfolios: one using the Sharpe maximization methodology, and the other using the Kurtosis minimization methodology.

We initially estimated the input parameters required for computing optimal portfolio weights according to the two strategies using a T -period-long dataset of returns. We then used subsequent periods, expressed in weeks, to assess the portfolio performance over the following W periods. The value of W was determined by the chosen rebalancing frequency. Throughout these W periods, we implemented a buy-and-hold approach, which is more practical for real-world investors compared to a constant mixed approach.

To evaluate and optimize the portfolios, we followed an iterative process. This process involved shifting the estimation window forward by a certain number of periods (W) and then recalculating the optimal portfolio weights. We repeated this process for the subsequent W periods while discarding an equal number of earliest returns. This iterative approach helped ensuring a comprehensive evaluation and optimization of the portfolios. We continued this process until we reached the end of the dataset.

3.2. Sharpe Maximization Methodology Portfolio

Constructing a portfolio that maximizes the Sharpe ratio offers numerous advantages within portfolio management. Investors utilize the Sharpe ratio, a crucial metric for evaluating investment performance, as it assesses the risk-adjusted return of a portfolio. According to a study by Siegel and Woodgate (2007), under appropriate conditions, portfolios that maximize the Sharpe ratio tend to become more diversified, which indicates that maximizing the Sharpe ratio can enhance portfolio diversification.

Portfolio analysts widely employ the Sharpe maximization methodology in portfolio analysis, as it assesses the opportunity cost of investing in a particular portfolio relative to the Risk-Free Asset. In this study, we utilized the 28-day T-bill as a reference. To implement this methodology, we used rolling averages to smoothen the highly volatile database of prices and returns. We examined this smoothing process based on the period chosen for each strategy.

$$Sharpe = \frac{E(R_p) - R_F}{\sigma_p} \quad (2)$$

where $E(R_p)$ represents the expected return of the portfolio, R_F denotes the risk-free rate, and σ_p indicates standard deviation of the portfolio shown as $\sigma_p = \sqrt{\text{Variance of the Portfolio}}$.

We adjusted the annual risk-free rate to match the frequency of the rebalancing period, ensuring consistency and accuracy in our calculations. It is crucial to emphasize that whenever a negative Sharpe ratio emerged, meaning that the return of the Risk-Free Asset surpassed that of the portfolio in that specific period, we chose to use the naive portfolio weights. In other words, we conducted rebalancing in every period without exception.

3.3. Kurtosis Minimization Methodology Portfolio

The Kurtosis minimization methodology utilizes a descriptive statistic known as kurtosis to measure how data disperse between a distribution's center and tails. Kurtosis values suggest that a data distribution may have "heavy" tails, meaning they are thickly concentrated with observations or have long tails with extreme observations (Green et al. 2023). The kurtosis minimization methodology in Pearson's kurtosis model aims to regulate the kurtosis of the portfolio's assets by assigning lower weights to assets with higher kurtosis and vice versa. We performed this calculation during each rebalancing period to ensure we achieved the lowest possible kurtosis whenever we traded currencies.

$$kurtosis_{x_k} = \frac{\sum_i^n (x_{ki} - \bar{x}_k)^4}{n\sigma_{x_k}^4} \quad (3)$$

where $kurtosis_{x_k}$ is the kurtosis of the variable x_k , a normally distributed sample will have a kurtosis close to 3.

Like the Sharpe maximization methodology, the same considerations and calculations were applied in this methodology. Rolling averages were utilized to smooth the highly volatile database of prices and returns, with the smoothing period determined by the strategy's timeframe. These rolling averages were then used to determine the portfolio weights that would result in the lowest possible kurtosis in each period. Subsequently, these weights were utilized in the subsequent holding period. Formula (5) ensures that the asset with the highest kurtosis receives the least weight of investment, while assets with lower kurtosis are assigned higher weights, thereby optimizing the portfolio for kurtosis minimization.

The formula for calculating the Kurtosis Minimization Portfolio Return is as follows:

$$Kurtosis \text{ Minimization Portfolio Return} = \sum_{n=1}^i W_{Kn(t)} \times R_{n(t+1)} \quad (4)$$

where $W_{Sn(t)}$ represents the weight of asset n calculated with Kurtosis minimization in period t. $R_{i(t+1)}$ denotes the return of asset n in period t + 1.

$$w_{x_k} = \frac{\frac{1}{kurtosis_{x_k}}}{\sum \frac{1}{kurtosis_x}} \times 100\% \quad (5)$$

where $\sum \frac{1}{kurtosis_x}$ represents the sum of inverses of the kurtosis of all the variables in a portfolio.

4. Results and Discussion

We conducted an analysis using the buy-and-hold strategy across five distinct rebalancing frequencies: half a week, one week, two weeks, four weeks, and eight weeks. We complemented this analysis by applying three portfolio optimization strategies: creating a Naive portfolio, maximizing the Sharpe ratio, and minimizing kurtosis. Given the high volatility inherent in cryptocurrencies, we opted for shorter holding periods. These shorter intervals allow for more frequent rebalancing and are suitable for investors with a shorter investment horizon. The rationale behind this approach is to capitalize on potential returns within a condensed timeframe. Table 2 provides a comprehensive overview of the returns obtained under different strategies and methodologies for the holding period spanning from 2020 to 2023.

Table 2. The holding period returns from 2020 to 2023.

Rebalancing Frequency	Naive Portfolio	Sharpe Maximization Methodology	Kurtosis-Minimization Methodology
Half a week	713.14%	762.73%	1166.73%
One week	651.20%	638.98%	823.24%
Two weeks	643.05%	354.48%	650.53%
Four weeks	386.66%	175.22%	682.84%
Eight weeks	293.93%	1225.68%	628.36%

Source: elaborated by the author. The figures in bold indicate the highest return achieved under either the Naive strategy, Sharpe maximization strategy, or kurtosis minimization strategy.

We further investigated the effectiveness of the optimization strategies by reducing the holding period to one year and evaluating their performance under the same rebalancing frequencies. Table 3 presents the returns achieved with different rebalancing strategies over a one-year time frame. Through this analysis, we aimed to determine which optimization strategy proves superior in optimizing returns within this extended investment horizon while maintaining consistent rebalancing frequencies.

Table 3. The holding period returns for one year.

Rebalancing Frequency	Naive Portfolio	Sharpe Maximization Methodology	Kurtosis Minimization Methodology
Half a week	90.62%	94.13%	118.50%
One week	86.09%	85.10%	98.23%
Two weeks	76.18%	59.37%	85.98%
Four weeks	62.76%	36.57%	88.41%
Eight weeks	52.51%	121.58%	84.28%

Source: elaborated by the author. The figures in bold indicate the highest return achieved under either the Naive strategy, Sharpe maximization strategy, or kurtosis minimization strategy.

The findings from Tables 2 and 3 indicate that the portfolio optimization method based on kurtosis minimization consistently outperformed others in the short term, particularly over holding periods of half a week, one week, two weeks, and four weeks, delivering optimal returns to investors. This observation aligns with the nature of cryptocurrencies, which are often traded in shorter terms, making kurtosis minimization particularly beneficial for capturing their market dynamics. This finding is corroborated by the literature, which emphasizes the importance of employing short-term strategies across various durations to construct comprehensive portfolios in cryptocurrencies. Each of these strategies serves a

distinct purpose, enabling investors to tailor their approaches to different market conditions and optimize risk-adjusted returns.

Figure 1 illustrates the distribution of weights along the estimation window when implementing the Sharpe maximization strategy. Most of the assets receive a portion of the investment throughout the entire period. Although there are sporadic periods where certain cryptocurrencies, mainly Ripple, did not contribute to the portfolio return, the weight allocation remained distributed among the ten assets.

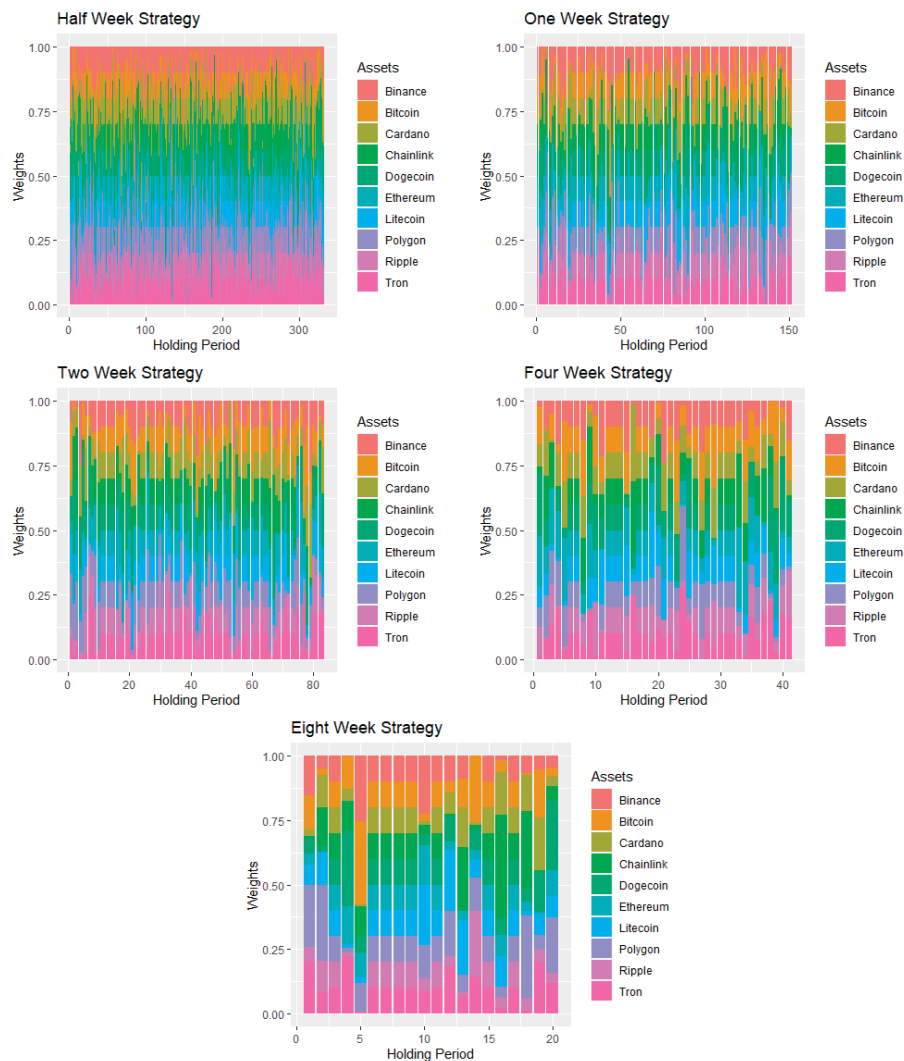


Figure 1. Bar charts representing the distribution of weights during holding periods with various rebalancing frequencies using the Sharpe ratio minimization strategy. Source: elaborated by the author.

Figure 2 depicts the distribution of weights during both holding and rebalancing periods within the kurtosis minimization strategy. There was no significant imbalance observed across any of the depicted strategies. However, a notable observation is the presence of a white line in the figures corresponding to the half-, one-, and two-week

periods. This occurrence in April 2022 occurred because kurtosis was not calculated for the portfolio due to the absence of movement in Litecoin’s price.

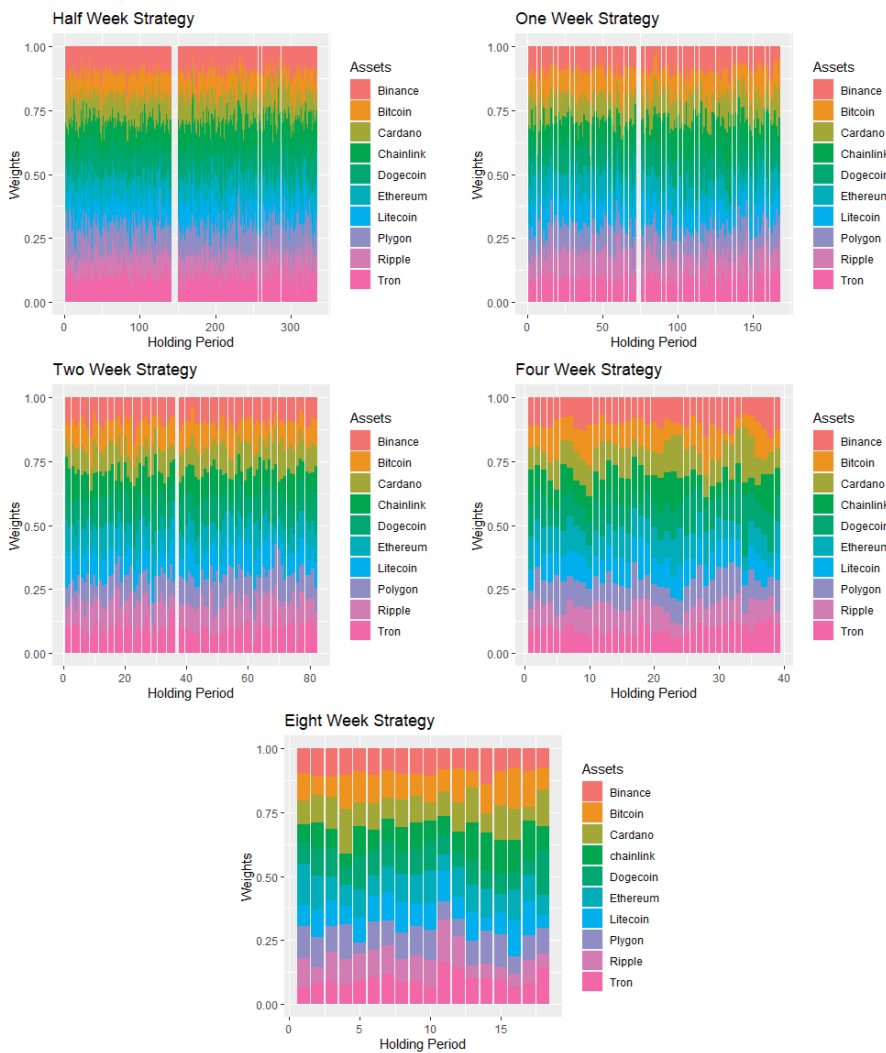


Figure 2. Bar charts representing the distribution of weights during holding and rebalancing periods using the kurtosis minimization strategy. Source: elaborated by the author.

Comparing the two methodologies, while the Sharpe maximization approach generated a more random distribution of portfolio weights along the estimation window, the kurtosis minimization method presented a steadier and more homogeneous distribution. Graphically, a consistent pattern is evident in the kurtosis minimization methodology, with most cryptocurrencies appearing in every rebalancing period. Conversely, the Sharpe maximization methodology exhibits a less patterned distribution, with several periods where some assets did not receive any investment.

Figure 3 presents a comparison of how portfolio returns behave under each methodology, segmented by strategy, revealing distinct patterns. The graphs illustrate that both the kurtosis minimization and Sharpe maximization methodologies effectively reduced the impact of high volatility on portfolio returns, and this was especially noticeable during

periods of heightened volatility, such as those observed in the naive portfolio strategy. Furthermore, the graphs show that strategies with higher rebalancing frequencies experience more pronounced reductions in volatility.

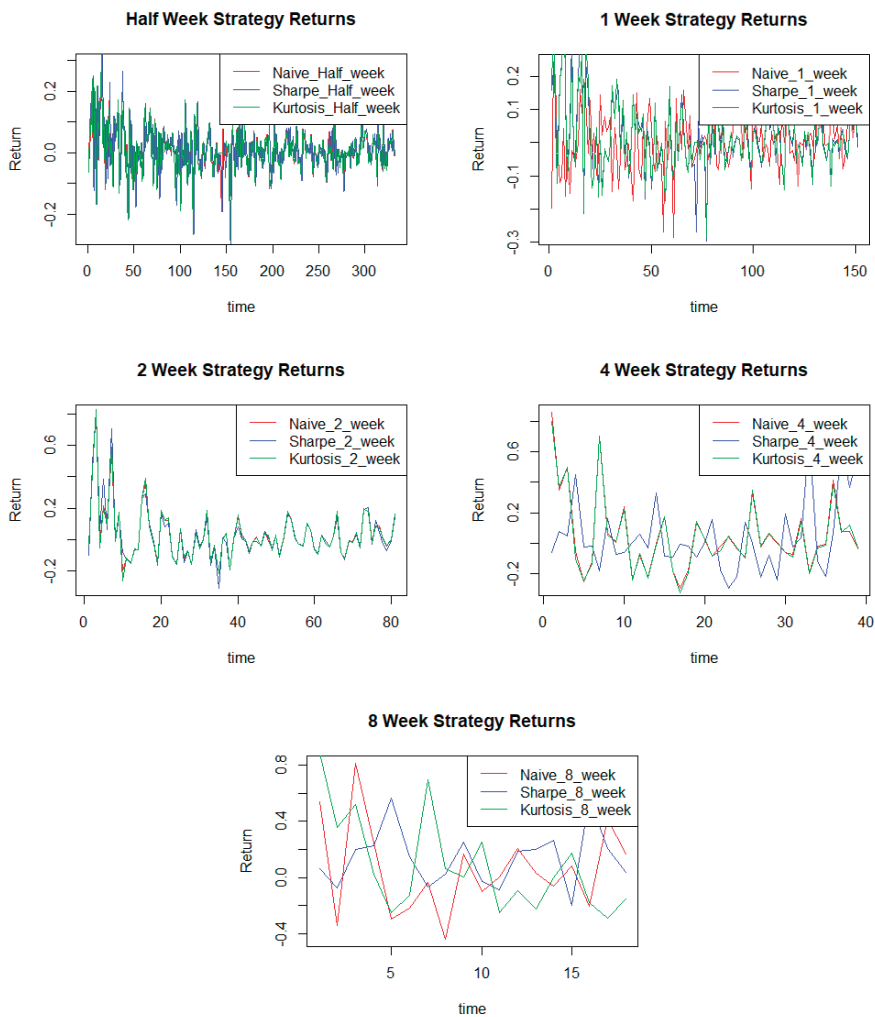


Figure 3. Comparison of portfolio returns behavior under different methodologies and strategies. Source: elaborated by the author.

In contrast to the findings of Ma et al. (2020), who emphasized the superior performance of the naive portfolio using the Sharpe ratio across various cryptocurrency portfolios in terms of return and risk profile, our study revealed the effectiveness of the kurtosis minimization portfolio, particularly within shorter-term rebalancing frequencies. While Ma et al. focused on longer-term diversification strategies, our research indicates that kurtosis minimization optimization yields favorable outcomes even in shorter investment horizons.

Similarly, Letho et al. (2022) underscored the diversification benefits of cryptocurrencies within an emerging market economy investment portfolio, employing metrics such as mean-variance analysis and the Sharpe ratio to support their findings. Our study aligns with this perspective, demonstrating that cryptocurrencies indeed offer diversification benefits. Moreover, leveraging the kurtosis minimization optimization can enhance portfolio

design by mitigating the risks associated with outliers and extreme market conditions, as supported by the findings of Cesarone et al. (2021).

Contrary to the empirical study conducted by Schellinger (2020), which prioritized maximizing the Sharpe ratio for optimizing cryptocurrency portfolios, our research challenges this approach by advocating for kurtosis minimization as a superior methodology, especially for short-term rebalancing frequencies. While Schellinger extended the traditional mean-variance approach, our findings suggest that focusing on minimizing kurtosis leads to more favorable results in cryptocurrency portfolio management.

The practical implications of our findings for portfolio managers and investors in the cryptocurrency market are significant and diverse. Our analysis highlights the necessity of employing advanced portfolio optimization techniques tailored to the unique attributes of cryptocurrencies. Specifically, we emphasize the effectiveness of the kurtosis minimization methodology in enhancing portfolio performance, particularly within shorter-term investment horizons. This insight equips portfolio managers with a valuable tool for maximizing returns while simultaneously mitigating risk in their cryptocurrency holdings.

Kurtosis minimization aims to diminish the impact of extreme returns or outliers in a portfolio, thereby enhancing its resilience to unforeseen market events and volatility (Wang et al. 2022). Conversely, Sharpe ratio maximization focuses on optimizing the risk-adjusted return of a portfolio by balancing risk and return to achieve higher returns for a given level of risk (Kumar and Shahid 2023). In scenarios where investors aim to mitigate potential risks associated with extreme market movements, kurtosis minimization is particularly relevant, aiming to create a stable and robust portfolio. By minimizing kurtosis, investors can reduce the likelihood of significant losses during turbulent market conditions, thereby improving overall risk management.

Furthermore, our research emphasizes the importance of regular portfolio rebalancing, especially in the volatile cryptocurrency market. By adopting a dynamic approach to portfolio management, investors can adapt to changing market conditions and optimize their portfolios for maximum returns. The use of shorter rebalancing frequencies, as advocated in our study, enables investors to capitalize on short-term opportunities and adjust their portfolios accordingly.

Regarding risk management strategies, our analysis underscores the significance of controlling factors such as kurtosis in portfolio construction. By minimizing kurtosis, investors can reduce the likelihood of extreme market events and enhance portfolio stability. This risk management approach is particularly relevant in the cryptocurrency market, characterized by inherent volatility and significant consequences of unexpected price movements.

5. Conclusions

The emergence of cryptocurrencies has brought about unique opportunities and challenges for investors and portfolio managers. In this study, we examined various portfolio optimization techniques and short-term investment strategies in the context of the cryptocurrency market, focusing on ten major cryptocurrencies.

Our analysis revealed several important findings. Firstly, we observed that cryptocurrencies are highly volatile and exhibit significant skewness and kurtosis. This highlights the need for sophisticated portfolio management techniques that are tailored to their unique characteristics. Secondly, we found that the kurtosis minimization methodology consistently outperformed other optimization strategies, delivering optimal returns to investors, especially in shorter-term investment horizons. This underscores the importance of controlling factors such as kurtosis for effective risk management and portfolio stability in the cryptocurrency market.

Our research also emphasized the importance of regular portfolio rebalancing, particularly in the volatile cryptocurrency market. A dynamic approach to portfolio management allows investors to take advantage of short-term opportunities and optimize returns.

Shorter rebalancing frequencies enables investors to adapt quickly to changing market conditions, thereby maximizing returns and mitigating risks effectively.

In conclusion, the insights gained from our analysis can guide decision-making processes, risk management strategies, and portfolio construction techniques in the cryptocurrency market. By incorporating the kurtosis minimization methodology, diversifying portfolios with cryptocurrencies, adopting regular rebalancing practices, and prioritizing risk management, portfolio managers and investors can navigate the complexities of the cryptocurrency market more effectively and optimize their returns.

Author Contributions: Conceptualization, S.S.; methodology, S.S. and J.H.O.V.; software, J.H.O.V.; validation, S.S., A.F.R. and J.-M.K.; formal analysis, S.S.; investigation, S.S.; resources, S.S. and J.H.O.V.; data curation, J.H.O.V.; writing—original draft preparation, S.S.; writing—review and editing, S.S., J.H.O.V. and A.F.R.; visualization, S.S.; supervision, J.-M.K.; project administration, S.S.; funding acquisition, J.-M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used were downloaded from the publicly available data on <https://www.cryptodatadownload.com/data/gemini/>, (accessed on 9 March 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Adrianus, Rio, and Subiaktio Soekarno. 2018. Determinants of momentum strategy and return in short time horizon: Case in Indonesian stock market. *International Journal of Trade and Global Markets* 11: 1. [CrossRef]
- Akyildirim, Erdinc, Ahmet Goncu, and Ahmet Sensoy. 2021. Prediction of cryptocurrency returns using machine learning. *Annals of Operations Research* 297: 3–36. [CrossRef]
- Alipour, Peyman, and Sina E. Charandabi. 2023. Analyzing the interaction between tweet sentiments and price volatility of cryptocurrencies. *European Journal of Business Management and Research* 8: 211–15. [CrossRef]
- Araujo, Fernando Henrique Antunes, Leonardo H. S. Fernandes, José W. L. Silva, Kleber E. S. Sobrinho, and Benjamin Tabak. 2023. Assessment the predictability in the price dynamics for the top 10 cryptocurrencies: The impacts of Russia–Ukraine war. *Fractals* 31. [CrossRef]
- Bagus, Philipp, and Luis P. de la Horra. 2021. An ethical defense of cryptocurrencies. *Business Ethics the Environment & Responsibility* 30: 423–31. [CrossRef]
- Ballis, Antonis, and Thanos Verousis. 2022. Behavioural finance and cryptocurrencies. *Review of Behavioral Finance* 14: 545–62. [CrossRef]
- Butterfill, J. 2024. Digital Asset Fund Flows | March 11. Etp.coinshares.com. Available online: <https://etp.coinshares.com/knowledge/market-activity/fund-flows-11-03-24> (accessed on 11 March 2024).
- Cesarone, Francesco, Manuel L. Martino, and Fabio Tardella. 2021. Mean-variance-var portfolios: Miqp formulation and performance analysis. *OR Spectrum* 45: 1043–69. [CrossRef]
- Chen, Yinnan, Xinchao Zhao, and Jianmei Yuan. 2022. Swarm intelligence algorithms for portfolio optimization problems: Overview and recent advances. *Mobile Information Systems* 2022: 1–15. [CrossRef]
- Choi, Sooyeon, and Richard A. Feinberg. 2022. Linking consumer innovativeness to the cryptocurrency intention: Moderating effect of the lohas (lifestyle of health and sustainability) lifestyle. *Journal of Engineering Research and Sciences* 1: 1–8. [CrossRef]
- CoinMarketCap. 2024. Cryptocurrency Market Capitalizations | CoinMarketCap. CoinMarketCap. Available online: <https://coinmarketcap.com/> (accessed on 11 March 2024).
- Conlon, Thomas, Shaen Corbet, and Richard J. McGee. 2020. Are cryptocurrencies a safe haven for equity markets? an international perspective from the COVID-19 pandemic. *Research in International Business and Finance* 54: 101248. [CrossRef]
- Dabbous, Amal, May Merhej Sayegh, and Karine Aoun Barakat. 2022. Understanding the adoption of cryptocurrencies for financial transactions within a high-risk context. *The Journal of Risk Finance* 23: 349–67. [CrossRef]
- ElBahrawy, Abeer, Laura Alessandretti, Anne Kandler, Romualdo Pastor-Satorras, and Andrea Baronchelli. 2017. Evolutionary dynamics of the cryptocurrency market. *Royal Society Open Science* 4: 170623. [CrossRef]
- Green, Jennifer L., Sarah E. Manski, Timothy A. Hansen, and Jennifer E. Broatch. 2023. Descriptive statistics. In *International Encyclopedia of Education*, 4th ed. Edited by Robert J. Tierney, Fazal Rizvi and Kadriye Ercikan. Amsterdam: Elsevier, pp. 723–33, ISBN 9780128186299. [CrossRef]
- Gu, Qianyu. 2022. Comparison of risk minimizing and return maximizing portfolio models. *BCP Business & Management* 30: 489–95. [CrossRef]
- Gupta, Nupur, Pradip Mitra, and Debjani Banerjee. 2022. Cryptocurrencies and traditional assets: Decoding the analogy from emerging economies with crypto usage. *Investment Management and Financial Innovations* 20: 1–13. [CrossRef]
- Hanif, Waqas, Hee-Un Ko, Linh Pham, and Sang Hoon Kang. 2023. Dynamic connectedness and network in the high moments of cryptocurrency, stock, and commodity markets. *Financial Innovation* 9: 1–40. [CrossRef]

- Inci, A. Can, and Rachel Lagasse. 2019. Cryptocurrencies: Applications and investment opportunities. *Journal of Capital Markets Studies* 3: 98–112. [CrossRef]
- Jana, Susovon, Ankita Nandi, and Tarak Nath Sahu. 2024. Can cryptocurrencies provide better diversification benefits? evidence from the indian stock market. *Journal of Interdisciplinary Economics* 15: 551–73. [CrossRef]
- Khan, Saleh Nawaz. 2022. The legality of cryptocurrency from an islamic perspective: A research note. *Journal of Islamic Accounting and Business Research* 14: 289–94. [CrossRef]
- Kong, Silin, and Mingchen Xu. 2023. Portfolio optimization for junior investors under different industries. *BCP Business & Management* 38: 1506–15. [CrossRef]
- Kozlovskiy, Serhii, Iaroslav Petrunenko, Hennadii Mazur, Vira Butenko, and Najatalya Ivanyuta. 2022. Assessing the probability of bankruptcy when investing in cryptocurrency. *Investment Management and Financial Innovations* 19: 312–21. [CrossRef]
- Krishna, Jaya Shankar, Akankhya Panda, and Pappu Sindhuja. 2023. Cryptocurrency adoption and its influence on traditional financial markets. *Interantional Journal of Scientific Research in Engineering and Management* 7: 1–11. [CrossRef]
- Kumar, Akhilesh, and Mohammad Shahid. 2023. Portfolio selection model using teaching learning-based optimization approach. *IAES International Journal of Artificial Intelligence (IJ-AI)* 12: 1083. [CrossRef]
- Letho, Lehlohonolo, Grieve Chelwa, and Abdul Latif Alhassan. 2022. Cryptocurrencies and portfolio diversification in an emerging market. *China Finance Review International* 12: 20–50. [CrossRef]
- Liu, Yongyu. 2022. Portfolio optimization with or without safe asset. Paper presented at 2022 International Conference on Economics, Smart Finance and Contemporary Trade (ESFCT 2022), Xi'an, China, July 22–24; pp. 425–32. [CrossRef]
- Liu, Yukun, Aleh Tsyvinski, and Xi Wu. 2022. Common risk factors in cryptocurrency. *The Journal of Finance* 77: 1133–77. [CrossRef]
- Lorenzo, Luis, and Javier Arroyo. 2023. Online risk-based portfolio allocation on subsets of crypto assets applying a prototype-based clustering algorithm. *Financial Innovation* 9: 1–40. [CrossRef]
- Ma, Yechi, Ferhana Ahmad, Miao Liu, and Zilong Wang. 2020. Portfolio optimization in the era of digital financialization using cryptocurrencies. *Technological Forecasting and Social Change* 161: 120265. [CrossRef]
- McMorrow, Jake, and Mona Seyed Esfahani. 2021. An exploration into people's perception and intention on using cryptocurrencies. *Holistica—Journal of Business and Public Administration* 12: 109–44. [CrossRef]
- Mungo, Luca, Silvia Bartolucci, and Laura Alessandretti. 2023. Cryptocurrency co-investment network: Token returns reflect investment patterns. *EPJ Data Science* 13: 11. [CrossRef]
- Nasir, Nurul Afaaf Mohd, and Hishamuddin Abdul Wahab. 2022. Analysing time-scales currency exposure using maximal overlap discrete wavelet transform (modwt) for malaysian industrial. Paper presented at International Conference on Mathematical Sciences and Statistics 2022 (ICMSS 2022), Online, March 15–16; pp. 426–39. [CrossRef]
- Petukhina, Alla A., Raphael C. G. Reule, and Wolfgang Karl Härdle. 2020. Rise of the machines? Intraday high-frequency trading patterns of cryptocurrencies. *European Journal of Finance* 27: 8–30. [CrossRef]
- Plerou, Vasiliki, Parameswaran Gopikrishnan, Bernd Rosenow, Luis A. Nunes Amaral, Thomas Guhr, and H. Eugene Stanley. 2002. Random matrix approach to cross correlations in financial data. *Physical Review E* 65: 066126. [CrossRef]
- Rajharia, Poonam. 2023. Cryptocurrency adoption and its implications: A literature review. *E3s Web of Conferences* 456: 03002. [CrossRef]
- Ruban, Aleksei Aleksandrovich, Alina Viktorovna Krivopuskova, and Dmitrii Sergeevich Kleimenov. 2022. Анализ эффектов включения биткойна в состав инвестиционных портфелей. *финансы и управление* 61–78. [CrossRef]
- Scagliarini, Tomas, Giuseppe Pappalardo, Alessio Emanuele Biondo, Alessandro Pluchino, Andrea Rapisarda, and Sebastiano Stramaglia. 2022. Pairwise and high-order dependencies in the cryptocurrency trading network. *Scientific Reports* 12: 1–20. [CrossRef] [PubMed]
- Schellinger, Benjamin. 2020. Optimization of special cryptocurrency portfolios. *The Journal of Risk Finance* 21: 127–57. [CrossRef]
- Shahzad, Muhammad. 2024. Cryptocurrency awareness, acceptance, and adoption: The role of trust as a cornerstone. *Humanities and Social Sciences Communications* 11: 1–14. [CrossRef]
- Siegel, Andrew F., and Artemiza Woodgate. 2007. Performance of portfolios optimized with estimation error. *Management Science* 53: 1005–15. [CrossRef]
- Silva, Isabela Ruiz Roque, Eli Hadad Junior, and Pedro Paulo Balbi. 2022. Cryptocurrencies trading algorithms: A review. *Journal of Forecasting* 41: 1661–68. [CrossRef]
- Subramoney, Stephanie Danielle, Knowledge Chinhamu, and Retius Chifurira. 2023. Var estimation using extreme value mixture models for cryptocurrencies. *Preprints* 2023: 2023070187. [CrossRef]
- Sukumaran, Shangeetha, Thai Siew Bee, and Shaista Wasiuzzaman. 2022. Cryptocurrency as an investment: The malaysian context. *Risks* 10: 86. [CrossRef]
- Susilo, Didik, Sugeng Wahyudi, Irene Rini Demi Pangestuti, Bayu Adi Nugroho, and Robiyanto Robiyanto. 2020. Cryptocurrencies: Hedging opportunities from domestic perspectives in southeast asia emerging markets. *Sage Open* 10: 215824402097160. [CrossRef]
- Symitsi, Efthymia, and Konstantinos J. Chalvatzis. 2019. The economic value of bitcoin: A portfolio analysis of currencies, gold, oil and stocks. *Research in International Business and Finance* 48: 97–110. [CrossRef]
- Trimborn, Simon, Mingyang Li, and Wolfgang Karl Härdle. 2019. Investing with cryptocurrencies—A liquidity constrained investment approach. *Journal of Financial Econometrics* 18: 280–306. [CrossRef]

- Wang, Yanqing. 2024. Do cryptocurrency investors in the uk need more protection? *Journal of Financial Regulation and Compliance*. [CrossRef]
- Wang, Zhuocheng, Zhiwei Huang, Rongkuan He, and Yixin Feng. 2022. *Cryptocurrency and the Herd Behavior*. Amsterdam: IOS Press. [CrossRef]
- Waspada, Ikaputera, Dwi Fitrizal Salim, and Astrie Krisnawati. 2022. Horizon of cryptocurrency before vs. during COVID-19. *Investment Management and Financial Innovations* 20: 14–25. [CrossRef]
- Yousaf, Imran, and Shoaib Ali. 2020. Discovering interlinkages between major cryptocurrencies using high-frequency data: New evidence from COVID-19 pandemic. *Financial Innovation* 6: 1–18. [CrossRef]
- Zaretta, Bara, and Irene Rini Demi Pangestuti. 2023. A methodological point of view in a systematic review of the documentation of the effect of volatility on cryptocurrency investment. *International Conference on Research and Development (ICORAD)* 2: 7–23. [CrossRef]
- Zhang, Mianmian. 2024. Relationships among return and liquidity of cryptocurrencies. *Financial Innovation* 10: 1–30. [CrossRef]
- Zhang, Shicheng. 2023. Portfolio management for multi-industry. *Highlights in Business, Economics and Management* 5: 214–21. [CrossRef]
- Zhang, Yuanyuan, Stephen Chan, Jeffrey Chu, and Hana Sulieman. 2020. On the market efficiency and liquidity of high-frequency cryptocurrencies in a bull and bear market. *Journal of Risk and Financial Management* 13: 8. [CrossRef]
- Zhao, Haidong, and Lini Zhang. 2021. Financial literacy or investment experience: Which is more influential in cryptocurrency investment? *The International Journal of Bank Marketing* 39: 1208–26. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

Segmenting Bitcoin Transactions for Price Movement Prediction

Yuxin Zhang ¹, Rajiv Garg ², Linda L. Golden ³, Patrick L. Brockett ^{4,*} and Ajit Sharma ¹

¹ Technology, Information Systems and Analytics, Mike Ilitch School of Business, Wayne State University, Detroit, MI 48201, USA; yuxin_zhang@wayne.edu (Y.Z.); gm5546@wayne.edu (A.S.)

² Information Systems and Operations Management, Goizueta Business School, Emory University, Atlanta, GA 30322, USA; rajiv.garg@emory.edu

³ Department of Marketing, McCombs School of Business, The University of Texas at Austin, Austin, TX 78712, USA; linda.golden@utexas.edu

⁴ Department of Information, Risk and Operations Management, McCombs School of Business, The University of Texas at Austin, Austin, TX 78712, USA

* Correspondence: brockett@utexas.edu

Abstract: Cryptocurrencies like Bitcoin have received substantial attention from financial exchanges. Unfortunately, arbitrage-based financial market price prediction models are ineffective for cryptocurrencies. In this paper, we utilize standard machine learning models and publicly available transaction data in blocks to predict the direction of Bitcoin price movement. We illustrate our methodology using data we merged from the Bitcoin blockchain and various online sources. This gave us the Bitcoin transaction history (block IDs, block timestamps, transaction IDs, senders' addresses, receivers' addresses, transaction amounts), as well as the market exchange price, for the period from 13 September 2011 to 5 May 2017. We show that segmenting publicly available transactions based on investor typology helps achieve higher prediction accuracy compared to the existing Bitcoin price movement prediction models in the literature. This transaction segmentation highlights the role of investor types in impacting financial markets. Managerially, the segmentation of financial transactions helps us understand the role of financial and cryptocurrency market participants in asset price movements. These findings provide further implications for risk management, financial regulation, and investment strategies in this new era of digital currencies.

Keywords: Bitcoin price direction prediction; blockchain transactions; financial transaction distribution; financial market segmentation; cryptocurrency transactions

Citation: Zhang, Yuxin, Rajiv Garg, Linda L. Golden, Patrick L. Brockett, and Ajit Sharma. 2024. Segmenting Bitcoin Transactions for Price Movement Prediction. *Journal of Risk and Financial Management* 17: 128. <https://doi.org/10.3390/jrfm17030128>

Academic Editor: Thanasis Stengos

Received: 8 February 2024

Revised: 6 March 2024

Accepted: 7 March 2024

Published: 21 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Cryptocurrencies have become an important component in the digital economy and financial markets (Adrian and Mancini-Griffoli 2021), with Bitcoin being the most popular. Bitcoin and blockchain technology are called the key drivers behind the next revolution in the financial industry (Li et al. 2021; Su et al. 2020). Despite Bitcoin's popularity, explosive growth, and innovation, it is often characterized as a financial bubble (e.g., Geuder et al. 2019; Cheah and Fry 2015). While the supply of Bitcoin is constrained, demand is speculative and causes wide fluctuations in price. Many governmental financial agencies, e.g., the U.S. Securities and Exchange Commission (SEC 2014), have issued investor alerts about Bitcoin due to the high risk associated with its price volatility (Aalborg et al. 2019). Studies of the extreme volatility in Bitcoin prices have also found evidence of both deliberate price manipulation Gandal et al. (2018) and substantial empirical support for the hypothesis that backers of the leading token Tether have the ability, and have possibly used Tether, to prop up the price of Bitcoin (Griffin and Shams 2020). Evidence of such market manipulation raises the specter of naive investors being left at the vagaries of an unregulated market manipulated by sophisticated, malicious actors engaging in deliberate and fraudulent price manipulation (Popper 2018). Investors can be unknowingly victimized without a

fundamental pricing method against which to benchmark. It is, therefore, crucial to better understand Bitcoin’s price movement and develop proper investment strategies, risk management, and financial regulation.

Several studies exist in which scholars have studied Bitcoin price movements (see Table 1). While these methods can help predict Bitcoin price movements to some extent, few have investigated the heterogeneous behavior of various investor segments and how investor heterogeneity influences Bitcoin price movement. In contrast, previous studies have shown that different types of investors react differently to stock price movements in the stock market, and this heterogeneity influences stock prices. Whether heterogeneity in investor types exists in the Bitcoin market and how this heterogeneity is related to Bitcoin price movement remains unclear and is one focus of this paper.

Table 1. Predicting the direction of the change in Bitcoin price.

Ref.	Data	Prediction Method	Prediction Period	Best Accuracy
Madan et al. (2014)	Historical Price	Generalized Linear Model	Next 10 min	0.57
Sattarov et al. (2020)	Twitter + Historical Price	Sentiment Analysis	Next 30 min	0.62
Sin and Wang (2017)	Historical Price	Artificial Neural Network	Next Day	0.60
McNally et al. (2018)	Historical Price	Long Short-Term Memory	Next Day	0.52
Kinderis et al. (2018)	Social Media (Twitter) + Historical Price	Linear Discriminant Analysis	Next Day	0.67
Atsalakis et al. (2019)	Historical Price	Artificial Neural Network	Next Day	0.67
		PATSOS Neuro-Fuzzy Controller Forecasting		
Arguelles (2018)	Historical Price	Support Vector Machine	Next Day	0.62
Kurbucz (2019)	Transaction Network	Single Hidden-Layer	Next Day	0.60
		Feedforward Neural Networks		

In this paper, we investigate investor segments in the Bitcoin market and study the heterogeneity in investment behavior. Based on our findings, we present a novel approach to predict the direction of Bitcoin price movements (up or down). Comparing our proposed approach with those in the extant research indicates the superior predictive power of our approach within a short investment duration. We also propose a typology of Bitcoin investors (i.e., investor segments) which, to the best of our knowledge, is the first attempt at systematically segmenting Bitcoin investors. This segmentation allows us to generate rich insights to help better understand the behaviors and actions of investors participating in Bitcoin markets. This information could also be utilized to better understand the role of investors’ participation in initial coin offerings and the short-term evolution of the coin (Holden and Malani 2022; Lyandres et al. 2022).

This paper is organized as follows. Section 2 reviews the current methods for predicting Bitcoin price movement and the challenges associated with them. Section 3 presents our investor segmentation approach and methodology. Section 4 presents the data analysis and results, and Section 5 interprets the results. Section 6 discusses the conclusions.

2. Neither Fish nor Fowl: A Review of Challenges in Current Predictions of Bitcoin Price Movements

The study of Bitcoin price movements is challenging for several reasons. First, as seen in Figure 1, Bitcoin prices exhibit very high daily volatility, making accurate prediction difficult (McNally et al. 2018; Miller and Kim 2021). Second, Bitcoin speculative assets are “neither fish nor fowl”—they are speculative investments that behave neither like traditional currencies nor usual physical commodities like wheat, oil, etc. (Gronwald 2019). This leads to higher fundamental uncertainties on the part of investors (Abraham et al. 2018). Extant research has shown that Bitcoin price volatility is highest at times when US

and European investors are not engaging in the trade of cryptocurrency assets (Dimpfl and Odelli 2020).

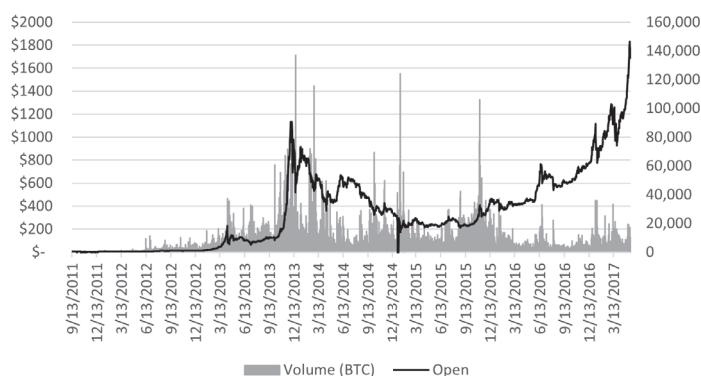


Figure 1. Bitcoin price volatility and daily transaction volume.

Economists generally say that “money” serves three functions: (1) it is a useful medium to facilitate the exchange of one product for another; (2) it should be able to serve as a unit of account so the comparative value or worth of different objects or entities can be measured and compared; and (3) it should provide a vehicle to store value. Currently, at best, Bitcoin only marginally satisfies these functions¹. The fact that it can take ten minutes to several hours (or more) to confirm that a Bitcoin transaction is valid makes using Bitcoin very inconvenient and problematic for everyday shopping transactions in daily life. The Lightning Network (LN) was invented in 2016 to allow for faster transaction verification, thereby better facilitating the use of Bitcoin for everyday transactions. The LN also has lower transaction fees. However, the LN comes with its own problems, such as the need to pre-establish a channel between the buyer and seller and show (post) liquidity to guarantee the validity of the transaction when opening an exchange arrangement. This may be problematic for small or infrequent purchases as it ties up resources. The LN also requires more technical skill to navigate.

Despite El Salvador’s adoption of Bitcoin as legal tender, the U.S. Internal Revenue Service considers Bitcoin assets to be property and not a currency for tax purposes (IRS Notice 2014-21, as modified by Notice 2023-34). The U.S. Commodity Futures Trading Commission (CFTC), which historically oversees physical commodity markets and commodity-linked products like derivatives and futures contracts, has classified the intangible Bitcoin asset as a commodity, although it does not derive its value from any physical asset such as gold, oil, etc. Thus, Bitcoin presents problems when used as a unit of exchange for most purposes (Alvarez et al. 2023).

Due to its extreme volatility, Bitcoin also presents problems in being perceived as a unit of account that could provide easy understanding and comparison of the value of goods and services relative to each other. As a store of value, the extreme volatility makes this problematic in the short run, and the intrinsic “financial bubble” aspects of this speculative asset, along with the potential for fraud or manipulation, make long-term storage of value problematic².

Thus, Bitcoin valuation is “free-floating”³ in the sense that its value is not benchmarked against or tied to other financial assets such as gold or silver. Neither is its value derived from its use as a raw material or intermediary resource in productive enterprises, e.g., rare-earth minerals have value because they are used in the manufacturing of permanent magnets and other high-tech products. Additionally, Bitcoin’s value is not based on interest rates and purchase price parity concepts, which are useful for currency exchange valuation. Furthermore, its value is not associated with production costs, e.g., computation costs and electrical power (Arnosti and Weinberg 2022). Some have argued that Bitcoin has intrinsic

value due to its portfolio diversification effect. Whether or not Bitcoin has a diversifying effect on portfolios, as with gold, is not clear. Some have observed diversifying effects (Li et al. 2021) and others have urged caution in using Bitcoin for portfolio diversification (Bakry et al. 2021). Almeida and Gonçalves (2023) provided a literature review of the portfolio diversification literature with respect to Bitcoin.

Consequently, economic theory and/or econometric arbitrage-free pricing methods, familiar for currency and asset pricing in finance, are missing. Finally, the relationships between Bitcoin market price, price volatility, and transaction behaviors are not yet well understood (Aalborg et al. 2019; Kjærland et al. 2018), making theoretical pricing models problematic. Moreover, due to random high price volatility, its use as a “storage of value” is uncertain at best. High volatility generates technical and economic risks for participants, system risks to the Bitcoin ecosystem, and societal risks at large, resembling a signal in high noise (Weaver 2018).

The absence of theoretically justified or arbitrage-based economic pricing models may not mean that predicting Bitcoin prices is futile. To assess the feasibility of obtaining a successful prediction of future prices, the finance concept of “weak form efficiency” is relevant. Applied to asset pricing, the weak-form efficient market hypothesis (EMH) asserts that publicly available information, such as an asset’s price history and volume history, cannot be used to make money in a highly liquid market by successfully predicting the asset’s future price direction because rationally informed investors will immediately incorporate past information into current prices in an unbiased manner. According to this theory, public information is already incorporated into current prices, and future prices become random fluctuations from today’s prices (the random-walk hypothesis). Some literature has tested the null hypothesis of weak-form efficiency for Bitcoin prices, examining whether prices follow a random walk, and has rejected this null hypothesis. Consequently, the EMH is not supported empirically for Bitcoin (e.g., see Bariviera 2017). In his literature review survey on efficiency in the cryptocurrency market, Kyriazis (2019) stated, “It is found that the majority of academic papers provide evidence for the inefficiency of Bitcoin and other digital currencies of primary importance.” Since prices may not follow a random walk, implying there may be dependence on Bitcoin price changes, it is worthwhile to examine whether public data and dependence on prices can be exploited for successful prediction of Bitcoin price change direction (cf., Fischer et al. 2019).

Apart from the EMH financial analysis, there is a burgeoning literature that technically addresses the prediction of Bitcoin price using various statistical and algorithmic methods, rather than methods derived from theoretical financial or economic relationships. For example, using ancillary information from social media platforms such as Twitter, News, and Google Trends, (Dutta et al. 2020). Mai et al. (2018) and Bartolucci et al. (2020) found that social media characteristics, such as the frequency of Bitcoin-related news or emotions in online discussions, can partially explain Bitcoin price movements. User sentiments being able to influence the price seems reasonable in light of the above discussion on Bitcoin pricing. Bitcoin price movement is not based on underlying economic fundamentals but is primarily speculative. The viral quality of social media discourse taps into, reflects, and even amplifies this speculative nature of cryptocurrency price determination.

A fundamental effort, therefore, concerns constructing quantitative models to predict Bitcoin prices based on historical price information. The first step in Bitcoin price prediction is to predict the direction (going up, staying stagnant, or going down) of Bitcoin price changes (e.g., should one buy or not). Table 1 highlights the research that has used a variety of machine-learning techniques and data sources for predicting the direction of Bitcoin price movements.

A review of the studies in Table 1 reveals that (1) most existing studies predicted daily Bitcoin price changes, (2) some studies predicted at the block time period using historical price data but performed poorly, and (3) two studies that used transaction information achieved relatively low accuracy scores. While a recent machine learning approach used historical (eight-month-old) data and a random forest model to achieve high prediction

accuracy of price movement (Amjad and Shah 2016), there is little evidence that such old data can causally influence the recent price of a highly volatile Bitcoin. Chen (2023) also applied machine learning (random forest regression and LSTM) to predict Bitcoin prices and found that higher accuracy for next-day price prediction was achieved when using only a one-period look-back on explanatory variables. In summary, the existing literature shows that when using historical price data, the predictive accuracy over a short time period is very challenging, whereas over a longer time period, accuracy is not significantly better than chance. Our study, however, sheds light on the predictability of price changes at a block level, within a short time frame. Our study not only achieves a higher prediction accuracy but also provides the possible mechanisms underlying Bitcoin price movement and the heterogeneity of investors.

While some existing studies have exploited Bitcoin transaction network data to predict Bitcoin price movement, a similarly important yet less explored topic is the converse: whether and how Bitcoin price affects Bitcoin transactions, or more specifically, how different segments of investors react differently to Bitcoin price movement. There are two questions worth investigating, which are discussed later in this paper. First, do changes in Bitcoin price, in turn, affect Bitcoin transaction behavior? And second, are such relationships heterogeneous, i.e., based on Bitcoin transaction characteristics? Some researchers have already found linkages between Bitcoin price and transaction network characteristics—such as connections among Bitcoin market participants (Akcora et al. 2018) and transaction volume (Koutmos 2018)—that are bidirectional and dynamic.

3. Not All Blocks Are Created Equal: Investor Segmentation

Each block in a Bitcoin blockchain contains all transactions that occur in approximately a rolling 10-min time window. The number of Bitcoin transactions occurring in each block, however, can vary significantly. Also, transactions can range in size from a fraction of a Bitcoin to a large number of Bitcoins. Because of this, Bitcoin transaction blocks are not uniform in composition and can differ in attributes such as the number and size of transactions. We further observe that Bitcoin blocks show interesting patterns when segmenting transactions in each block by transaction amounts. Figure 2 shows an example of how blocks evolve when transactions are segmented by transaction amount (in BTC).

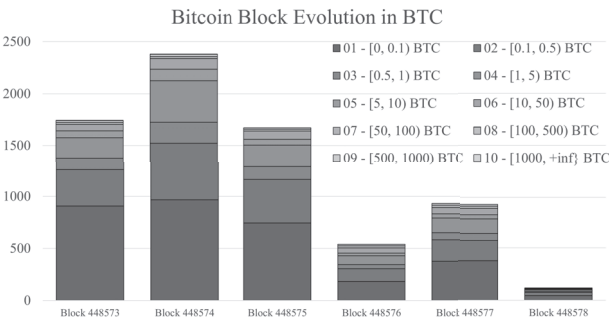


Figure 2. Number of transactions by block.

Figure 2 shows that the number of transactions in each transaction class varies significantly across blocks. Additionally, if we calculate the proportion of transactions falling into each transaction class, it also varies significantly by block. Therefore, given the high price volatility associated with the high volatility of the transaction classes, we propose that analogous to what is found in the stock market, there are multiple investor types in the Bitcoin market, and if classified by their transaction amount, different types of investors are dynamically associated with Bitcoin price movement. Later, in Section 5, we present possible labels, i.e., different types of investors, for these transaction classes and provide possible underlying mechanisms to explain the associations.

Based on the above proposal, we develop an effective predictive model that captures the heterogeneity in the association between the types of Bitcoin investors and Bitcoin price movement. In our predictive model, to predict the directionality of BTC price movements over time, we use the change in transaction class in two ways: (1) the change in the number of transactions in each transaction class from one block to the next, and (2) the change in the proportion of each transaction class to the total transaction numbers from one block to the next. This allows us to capture the heterogeneity in speculations by various investor types and utilize their prior time-period decisions to predict future price movements. This segmentation of investors provides rich insights for more accurate prediction—not just for capturing volatility but also for understanding the behavior and actions of investors participating in this market. Details of operationalizing this approach are described in Section 4.

4. Data Analysis and Results

4.1. Data

Historical Bitcoin transaction data were collected and processed from three data sources: (1) blockchain.com (accessed on 21 March 2018); (2) blockcypher.com (accessed on 21 March 2018); and (3) Bitcoinchart.com (accessed on 21 March 2018). The first two sources provide Bitcoin transaction history data, including block IDs, block timestamps, transaction IDs, senders’ addresses, receivers’ addresses, and transaction amounts. They contain 306,033,506 transactions in total, dating from 18:15:05 on 3 January 2009 to 03:56:50 on 21 March 2018. The third source provides the Bitcoin market transaction price for every five minutes, with the time windows for this data source being from 13:50:00 on 13 September 2011 to 00:00:00 on 14 May 2017. We merged these three datasets to obtain the Bitcoin transaction history with market exchange price for the time period from 13:50:00 on 13 September 2011 to 00:00:00 on 14 May 2017.

4.2. Transaction Segmentation

For each block, we segmented the transactions into 15 segments based on transaction volume, as shown in Table 2. Here, the “Transaction Amount Segment” represents the transaction amount interval (in BTCs) in which a transaction falls.

Table 2. Transaction segmentation used in predicting the direction of the change in Bitcoin price.

Transaction Amount Segment	Total # of Transactions	Mean # of Transactions per Block
(0, 0.1] BTCs	145,731,106	448
(0.1, 0.5] BTCs	38,978,755	120
(0.5, 1] BTCs	12,944,521	40
(1, 5] BTCs	18,690,349	57
(5, 10] BTCs	5,199,054	16
(10, 25] BTCs	5,071,494	16
(25, 50] BTCs	2,677,391	8
(50, 100] BTCs	1,592,056	5
(100, 200] BTCs	887,452	3
(200, 500] BTCs	758,835	2
(500, 1000] BTCs	226,882	1
(1000, 5000] BTCs	150,458	0.5
(5000, 10,000] BTCs	37,608	0.1
(10,000, 50,000] BTCs	18,042	0.05
(50,000, ∞] BTCs	529	0.002

4.3. Price Movement Prediction

The baseline framework for predicting the direction of price movement (moves up or not at time T)⁴ is described as follows:

$$m_T = f(\vec{m}_{T-1}, \Delta \vec{N}_{T-1}, F_T) \quad (1)$$

Here, m_T , the predicted binary variable, can take the value of one or zero, with one representing an upward Bitcoin market price movement from time $T - 1$ to T , and zero representing an unchanged or downward market price movement from time $T - 1$ to T . The function f represents a binary prediction classifier that utilizes the following predictive variables:

- (1) $\vec{m}_{T-1}$, a vector representing the direction of the historical Bitcoin price movement from time $T - p$ to $T - 1$. This vector of zeros and ones is denoted by:

$$\vec{m}_{T-1} = (m_{T-1}, m_{T-2}, \dots, m_{T-p}) \quad (2)$$

where m_{T-t} represents the Bitcoin price movement direction from time $T - (t + 1)$ to $T - t$.

- (2) $\Delta \vec{N}_{T-1}$, a vector representing the change in transaction volume of all transaction classes during each time period from time $T - p$ to $T - 1$. This vector is denoted by:

$$\Delta \vec{N}_{T-1} = (\Delta n_{i,t})_{(1 \leq i \leq 15, T-p \leq t \leq T-1)} \quad (3a)$$

i.e.,

$$\Delta \vec{N}_{T-1} = \begin{pmatrix} \Delta n_{1,T-1}, \dots, \Delta n_{15,T-1} \\ \vdots \\ \Delta n_{1,T-(p-1)}, \dots, \Delta n_{15,T-(p-1)} \\ \Delta n_{1,T-p}, \dots, \Delta n_{15,T-p} \end{pmatrix} \quad (3b)$$

- (3) A variable F_T , representing a fixed effect measured at time T . We consider two types of time-specific fixed effects: the day/night fixed effects and the month fixed effects. We tested our results with various values of the look-back period (p). For consistency and simplicity, and to save space, we present the subsequent results only for the case $p = 10$.

Here, to accommodate two possible models, $\Delta n_{i,t}$ represents either the change in the number of transactions within transaction class i or the change in the proportion of transactions falling in transaction class i from time $t - 1$ to t depending on whether we are focusing on the changes in the number or the proportion of changes, respectively.

As mentioned previously, in our investigation, $\Delta n_{i,t}$ is defined in two alternative ways: (1) $\Delta n_{i,t}$ measures the change in the number of transactions in transaction segment i from time $t - 1$ to t , and (2) $\Delta n_{i,t}$ represents the change in the proportion of transactions in segment i from time $t - 1$ to t . The first approach allows us to capture the change in the number of transactions over time, whereas the second approach allows us to capture the change in the distribution of transactions over time. Additionally, including a time-specific fixed effect enables the model to capture transaction times across different countries, and consequently, the differences in transaction patterns caused by the time/location differences.

Additionally, we estimated the performances of multiple combinations of variables for the prediction of price movement: transaction volume/distribution changes in each transaction interval, historical price, day/night fixed effect, and month fixed effect (Table 3). Using the transaction volume changes and the transaction distribution changes, we demonstrated a significant predictive improvement compared with existing studies, suggesting an important role of publicly accessible transaction data in Bitcoin price movement.

We present the results using two classifiers. The first is recurrent neural network (RNN) classifiers since they have been shown to have a strong ability to learn and model time-series data with non-linear and complex relationships. A special form of RNN, the Long Short-Term Memory (LSTM) method, allows information to persist and is capable of learning from the past while selectively remembering patterns for a longer duration of time. LSTM provides a good prediction of the direction of Bitcoin price movement. Specifically, we constructed a three-layer LTSM model, using categorical cross-entropy as the loss function.

Table 3. Direction of Bitcoin price movement prediction.

Direction of Bitcoin Price Movement Prediction Using Transaction Volume Changes						
Model	Hist. Price	Trans. Volume	Month F.E.	Day/Night F.E.	Accuracy	F-1 Score
LSTM	Yes	Yes	No	Yes	0.631	0.773
LSTM	Yes	Yes	Yes	Yes	0.631	0.768
Logistic Regression	No	Yes	No	No	0.599	0.663
Logistic Regression	Yes	Yes	Yes	Yes	0.596	0.652
Logistic Regression	Yes	No	Yes	Yes	0.532	0.585
Logistic Regression	Yes	No	No	No	0.496	0.497
Direction of Bitcoin Price Movement Prediction Using Transaction Proportion Distribution Changes						
Model	Hist. Price	Trans. Volume	Month F.E.	Day/Night F.E.	Accuracy	F-1 Score
LSTM	Yes	Yes	Yes	Yes	0.636	0.778
LSTM	Yes	Yes	No	Yes	0.635	0.777
Logistic Regression	Yes	Yes	Yes	Yes	0.538	0.586
Logistic Regression	Yes	Yes	No	Yes	0.535	0.592

Another classifier used with the same sets of variables is the logistic regression model. Model fixed effects are included for possible day /night and month fixed effects. In the logistic regression model, the prediction function takes the form of:

$$P(m_T = 1) = \frac{\exp(\beta_0 + \sum_p^{t=1} \beta_{1,t} m_{T-t} + \sum_p^{t=1} \sum_{i=1}^{15} \beta_{2,i,t} \Delta n_{i,T-t} + \beta_3 F_T)}{1 + \exp(\beta_0 + \sum_p^{t=1} \beta_{1,t} m_{T-t} + \sum_p^{t=1} \sum_{i=1}^{15} \beta_{2,i,t} \Delta n_{i,T-t} + \beta_3 F_T)}$$

(4)

In this equation, p is the number of look-back periods for the transaction volume or transaction volume distribution changes; β_0 denotes the intercept; and $\beta_{1,t(1 \leq t \leq p)}$, $\beta_{2,i,t(1 \leq i \leq 15, 1 \leq t \leq p)}$, and β_3 denote the coefficients of the historical price movements, changes in the number (or proportion, respectively) of transactions in each volume segment, and the fixed effects vectors.⁵

4.4. Results

First, we report the prediction results for the direction of price movement based on transaction volume changes as the key predictive variable while incorporating other variables in the predictive model. The best prediction model results are presented in Table 3 (upper) with an accuracy rate of 63.1%. The best-performing predictive models were (a) the recurrent neural network method (LSTM), using the historical price movement, historical changes in the transaction volume, and day /night fixed effects (i.e., not including the month fixed effects), and (b) the LSTM method, incorporating all the variables introduced earlier, including the month fixed effects. While both LSTM models exhibited the same accuracy (number of correct predictions divided by the number of predictions), the first LSTM model exhibited a slightly higher F1-score (the harmonic mean of the precision and the recall), indicating greater overall accuracy as measured with the precision and recall rates. Since all the models in Table 3 were run on the same dataset and same forecast horizon (unlike

the summary given in Table 1), this result clearly shows that incorporating transaction volume changes improved performance. For comparison with the methods in Table 1, the LSTM metrics for prediction using historical data and not incorporating the transaction data or fixed effects yielded an accuracy of 0.4941 and an F1-score of 0.5292. This compares to (see top two rows of Table 3) the accuracy of 0.631 and F1-score of 0.768–0.773 for the same analysis on the same dataset with the same prediction horizon but also including the transaction network data. The common dataset and prediction horizon in Table 3 make it more fair to compare the results of Table 1 with those of Table 3 with the incorporation of a transaction volume variable. The LSTM models performed better when the transaction volume was incorporated.

Since Table 3 (upper) shows that the inclusion of historical changes in transaction volume provided a more accurate classification of the direction of price movement, this raises the question of how much the inclusion of transaction volume changes contributes to the prediction. To further explore this, Table 3 (upper) revisits the previous logistic regression models from Table 1, comparing the results when excluding versus including the historical and transaction volume changes from the explanatory variables (logistic regression was among the top methods on our data and also one of the most used quantitative methods employed in the literature presented in Table 1). The results in the last line of Table 3 (upper) show that including the historical price movement and fixed effects without including the historical record of transaction volume changes led to significantly worse performances in classifying price movement direction. These results provide empirical evidence of the central role transaction network characteristics play in predicting Bitcoin price movement. The comparison is fair because all models used the same dataset and prediction horizon.

Next, we replicated the analysis of Table 3 (lower), but now used the changes in the proportion (instead of the number) of transaction segments as predictive variables (as discussed in after Equation (3b) in Section 4.3). With the inclusion of the changes in transaction volume distributions in the model, we observed (Table 3 (lower)) a slight improvement in predictive accuracy for the recurrent neural network method when also including the historical price movement, monthly fixed effects, and day/night fixed effects. The highest accuracy rate reached was 63.6%. By contrast, the logistic regression model performed significantly worse, even relative to the logistic regression model that included transaction volume changes as the predictive variable (compared with the previous table).

4.5. The Relationship between Price Movement and Transaction Volume in the Bitcoin Market

Next, we explored the causal effect of transaction volume changes in each transaction segment on the direction of Bitcoin price movement. While our previous predictive models achieved relatively high accuracy, these models are inherently black-box models and did not explicitly explain how the transaction volume was related to the direction of Bitcoin price movement. Thus, we applied the linear probability model to interpret the direction of Bitcoin price changes in a time-series framework (Battay et al. 2019). Specifically, we regressed the binary direction of Bitcoin price movement on a series of historical transaction volume changes in each transaction segment while incorporating fixed effects. The coefficients reflect the explanatory power of each transaction volume segment at the examined block time prior to the current block time on the current price movement. The interplay between the transaction volume and block time allows us to explore how the explanatory power varies based on transaction size and time. The rationale behind this estimation strategy is to differentiate the impacts of the volume of small-amount and large-amount historical transactions and if they have different degrees of effect on future price changes. Thus, we estimate the following linear probability model:

$$P(m_T = 1) = \alpha_0 + \sum_{15}^{t=1} \sum_{15} \alpha_{i,t} \Delta n_{i,T-t} \quad (5)$$

The dependent variable in Equation (5) is the binary variable indicating whether the Bitcoin price moves upward ($m_T = 1$) versus staying stagnant or moving downward ($m_T = 0$). This model is similar to the earlier predictive model (Equation (1)). Through this linear probability model, we can interpret how a change in the number of transactions in different transaction amount segments generates heterogeneous effects on Bitcoin price movement.

Based on the estimated coefficients (Table 4), we can observe that a transaction volume (or transaction distribution) that is older than 10 blocks in the past does not have a significant impact on the expected current Bitcoin price movement⁶. Furthermore, interestingly, these coefficients suggest that a unit change in “small-amount” transactions positively affects Bitcoin price movement, whereas a unit change in “large-amount” transactions negatively affects Bitcoin price movement.

Table 4. Price movement by the change in the number of transactions.

Segment	$t - 1$	$t - 2$	$t - 3$	$t - 4$	$t - 5$	$t - 6$
(0, 0.1)	$4.52 \times 10^{-5} ***$	$4.14 \times 10^{-5} ***$	$3.74 \times 10^{-5} ***$	$4.10 \times 10^{-5} ***$	$3.95 \times 10^{-5} ***$	$3.81 \times 10^{-5} ***$
(0.1, 0.5)	$1.98 \times 10^{-4} ***$	$1.44 \times 10^{-4} ***$	$6.00 \times 10^{-5} *$			
(0.5, 1)	$3.70 \times 10^{-4} ***$	$3.03 \times 10^{-4} ***$	$3.18 \times 10^{-4} ***$	$2.04 \times 10^{-4} **$		
(1, 5)	$2.67 \times 10^{-4} ***$	$3.15 \times 10^{-4} ***$	$2.64 \times 10^{-4} ***$	$2.27 \times 10^{-4} ***$	$2.66 \times 10^{-4} ***$	$2.07 \times 10^{-4} ***$
(5, 10)	$8.91 \times 10^{-4} ***$	$6.82 \times 10^{-4} ***$	$8.26 \times 10^{-4} ***$	$8.18 \times 10^{-4} ***$	$6.77 \times 10^{-4} ***$	$5.85 \times 10^{-4} ***$
(10, 25)	$3.63 \times 10^{-4} ***$	$2.32 \times 10^{-4} *$				
(500, 1000)	$-3.22 \times 10^{-4} *$	$-3.85 \times 10^{-4} *$				
(1000, 5000)					$-4.67 \times 10^{-4} *$	
(5000, 10,000)		$-8.39 \times 10^{-4} *$				$-7.49 \times 10^{-4} *$

Significant codes: *** 0.001; ** 0.01; * 0.05.

We now turn to the specific transaction amount segments. First, although all the coefficients for the first transaction amount interval, 0 to 0.1 Bitcoins, were significant, the transactions in this interval were likely to be transaction fees and were less related to the amount of the transaction. The significance here could suggest that the volume of micro-transactions is a proxy for overall transactions happening in the marketplace. For the next four intervals, 0.1 to 10 BTC, we found that an increase in transactions in the segment 0.1 to 1 BTC had a shorter-term positive effect on the Bitcoin price with no effect observed beyond $t - 4$, whereas the positive effect of larger Bitcoin transactions, from 1 to 10 BTC, persisted over a longer term. Small-transaction investors whose transactions fall in the 0.1 to 1 BTC interval are more likely to be driven by market sentiments and speculations that affect the price in the short term. Since a small amount of Bitcoin is more affordable, potential amateur or uncertain buyers can purchase, and this could lead to a significant increase in demand. A longer-term effect was observed on transactions in the range of 1 to 10 BTC. This is intuitive because a participant with a stronger speculation motive is likely to invest more in such a dynamic financial market Dyhrberg et al. (1995). Thus, mid-scale transactions ranging from 1 to 10 BTC were much larger and more frequent, leading to significant effects on Bitcoin price movement in the long run. This result is consistent with earlier findings that the persistence of Bitcoin price volatility is driven by massive middle-scale transactions (Scharnowski 2021). By way of contrast, transactions ranging from 500 to 10,000 BTCs exhibited a negative effect on pricing. This is possibly because many investors in this large-dollar range are institutional investors, and they can benefit from investing in products that are classified as overvalued by timing the correction of overpricing Edelen et al. (2016); Jang and Kang (2019). As a result, it could be an indication of an anticipated price drop when these large-dollar investors become more active in making Bitcoin transactions.

5. Interpretation and Discussion

Our findings provide interesting insights into the dynamic relationship between Bitcoin price and transaction distribution. The findings suggest that various segments of

Bitcoin market participants influence price movement in intervals as short as one block or as long as two hours. We identify three groups (small, mid, and large) of participants based on transaction amounts. We then analyze how different types of market participants can potentially respond to price dynamics in different ways, and how the divergent patterns of market response can lead to changes in the dynamics of price and transactions. This utilizes theory from social influence and network density to shed information on price movement.

5.1. Market Participants

A rationale consistent with our Bitcoin market analysis is that different market participants respond to Bitcoin price movement in different ways and, in turn, also generate different effects on Bitcoin price. While these insights are similar to market segmentation present in financial equity markets, it is of particular importance to study the relationship between Bitcoin transactions and price movement given Bitcoin's high volatility Blau (2017), its speculative nature, and the public accessibility of its transaction information. The transaction information also enables information transparency that is usually held by large financial institutions in traditional financial markets. This information can help future market participants make more rational decisions to optimize returns from Bitcoin markets. It is expected that Bitcoin price and transactions have a dynamic relationship similar to assets in the stock market (Gabaix et al. 2006). On the one hand, traders with small and large transaction amounts respond differentially to price movement in the Bitcoin market, whereas on the other hand, trading behaviors can reflect (and even cause) price movement. Thus, we present three groups of participants, as seen in Figure 3: small investors (e.g., algorithmic traders), mid-sized investors (e.g., individuals), and institutional investors (e.g., individual whales). Figure 3 presents a contour map, where the x-axis represents the natural logarithm of transaction amounts, the y-axis represents the natural logarithm of the number of transactions in each block, and the contour lines and shading represent the total number of blocks. In this figure, we can see that the transactions show significant patterns and can be segmented into three segments: large transaction volume but small transaction amount, medium transaction volume and medium transaction amount, and low transaction volume but large transaction amount. These correspond to the three investor groups, which we describe next.

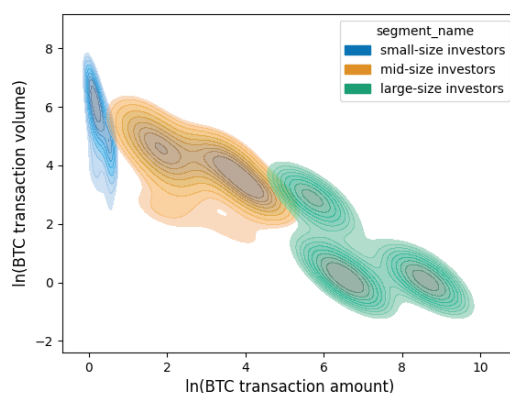


Figure 3. Bitcoin market participants and transaction patterns.

5.1.1. Individual Investors

Similar to individual participants in the stock market, individual Bitcoin investors mainly focus on small-to-medium transactions. In a financial market, individual market participants focusing on small-amount transactions actively adjust their investment strategies following changes in stock prices. For example, as discussed in earlier stock market studies (e.g., Niederhoffer and Osborne 1966), individual stock market investors are more

likely to invest when stock prices are increasing and less likely to invest when stock prices are decreasing. This type of investment behavior can be explained by behavioral financial theories (e.g., Ji et al. 2008), which argue that an individual's expectation of continued price increase in a "rising market" could surpass expectations of the risk associated with a price decline in the future. This transaction pattern is similar in the Bitcoin market, where we observed a long-term positive impact on price movement.

5.1.2. Algorithmic Bitcoin Traders (Micro-Traders)

Algorithmic traders concentrate on high-frequency transactions that lie within the smaller-size transaction segments. The goal of high-frequency transactions is fundamentally different from that of individual participants in the market Gsell (2008): algorithmic trading tends to focus on short-term costs and benefits, whereas individuals are relatively more interested in longer-term trends in the market price. Therefore, the optimal strategy of algorithmic trading is to capitalize on the benefits in a rising market by selling immediately after the price increases. In other words, in contrast to individual participants' investment behaviors, algorithmic trading tends toward "sell" as the price increases and "buy" as the price decreases. Both individual market participants and algorithm traders are sensitive to price movement, and a change in the market price leads to an increase in transactions for both types of participants with small transaction amounts in the Bitcoin market. This is consistent with our findings on the short-term positive relationship between small-size transaction volume and price movement.

For individual traders who concentrate on small transactions, a commonly exhibited trading strategy is to buy in a rising market and sell in a falling market Ji et al. (2008), whereas on the other hand, algorithmic trading tends toward buying in a falling market and selling in a rising market. In any case, market price movements generate a positive effect on transaction volume, with some small-amount market participants buying and others selling. It is thus possible to exploit price movement to predict the volume of small-amount transactions. Additionally, the buying and selling behaviors among small-amount participants could also reflect price movement, especially in the Bitcoin market where the supply of new Bitcoins is limited. Therefore, an increase in transactions marks a higher level of trading activity, suggesting a higher demand for Bitcoins in the market. This leads to a price rise as demand increases while the supply remains fairly stable. Although a causal link between small-amount transactions and a price rise or decline might be too strong an assumption for the Bitcoin market, any changes in the volume of small-amount transactions might still reflect—and thus can be used as a predictor of—price movement in the Bitcoin market.

5.1.3. Larger-Volume Traders (Institutions and Whales)

Typically, individual market participants and algorithm traders are not involved in large-scale transactions in the financial market due to their limited financial resources and capacity. Instead, market participants focusing on transactions with high amounts are generally large institutional traders who have sufficient financial resources to engage in such transactions (Chan and Lakonishok 1995). These investors likely possess private information on the long-term performance of an asset and tend to invest when the price is declining or sell when the price is increasing, which is again consistent with our findings regarding Bitcoin.

Institutional trading, which focuses on large-amount transactions, depends on the specific trading volume and the size of the investor (Gabaix et al. 2006). Given the high volatility of the Bitcoin market, institutional traders might be less inclined to engage in transactions due to risk aversion compared to individuals and algorithm traders (O'Connell and Teo 2009).

Therefore, there could be a divergent pattern of transactions under price movement between market participants who focus on small-amount transactions versus large-amount transactions. On the other hand, since institutional trading involving large-amount transac-

tions is disproportionately influential in the Bitcoin market (Aalborg et al. 2019), changes in the volume of large-amount transactions can generate a direct impact on the Bitcoin price, in addition to reflecting the price movement. The above differential dynamics provide a basis for predicting the mutual relationship between transaction volume and Bitcoin price in the upper tail of transactions.

5.2. Using Prediction of Price Direction Movement as an Investment Guide

To better illustrate the predictability and application of our model, we compared it to a very simple trading strategy based on a naive prediction model, the random-walk model. We then monitored the total asset value and ROI of each investment strategy at each block over a span of a month. The investment strategy used is simple: once block (number N) is mined, immediately invest the balance to purchase Bitcoins (BTCs) and hold it. Continue to hold it when the next block (number $N+1$) is mined, provided the predicted price movement from block N to $N+1$ is “up”. Alternatively, immediately sell all BTCs in the account if the predicted price movement from block N to $N+1$ is “stagnant or down”. We applied the same strategy based on our model prediction and a “random-walk” model prediction, which assumes that the Bitcoin price has a 50% chance of going up and a 50% chance of not going up.

The use of the random-walk model to provide a reasonable comparison model against our model may require a justification for some. Indeed, one might conjecture that it is easier for a machine learning model to outperform the random-walk model, especially in series with trends. However, for very short-term predictions (about ten minutes into the future for a block-length prediction), this conjecture is less obvious because of Bitcoin’s extremely high volatility. Additionally, there is empirical support for the random-walk hypothesis being a reasonable (non-rejectable) alternative model in some tests of the efficient market hypothesis for Bitcoin (e.g., Nadarajah and Chu 2017). Thus, we do not view the random-walk model as a “straw man” comparison but rather as a reasonably simple alternative model for comparison.

We can also find some support for using the random-walk model as an alternative model in the theoretical and empirical exchange rate literature. The paper by (Meese and Rogoff 1983) is an important and seminal paper in the currency exchange literature. It shows that for out-of-sample prediction, sophisticated exchange rate models in the literature do not perform better in short-term prediction (1–12-month horizons in their paper) compared to the random-walk model. According to Moosa and Burns (2014), “This view [that the random-walk model does as well as more sophisticated econometric finance models] is still widely accepted to the extent that it is typically argued that the Meese–Rogoff results, which are “yet to be overturned”, constitutes a puzzle.”

We include the random-walk model because it is a reasonable short-term alternative model, especially due to the high local volatility of Bitcoin prices. At any rate, in light of the currency literature cited above, as well as some literature that tests whether Bitcoin price changes are consistent with the weak-form (random walk) efficient market hypothesis, we feel that the random-walk model provides a reasonable comparison benchmark for short-term prediction.

The trading strategy based on our model reached a total return on investment (ROI) of 305%, whereas the naive random-walk model only reached a total ROI of 20%, with both strategies starting with the same initial capital investment. Figure 4 shows how the ROI changed over time based on these two strategies. The x-axis in Figure 4 represents the time, and the y-axis represents the ROI percentage.

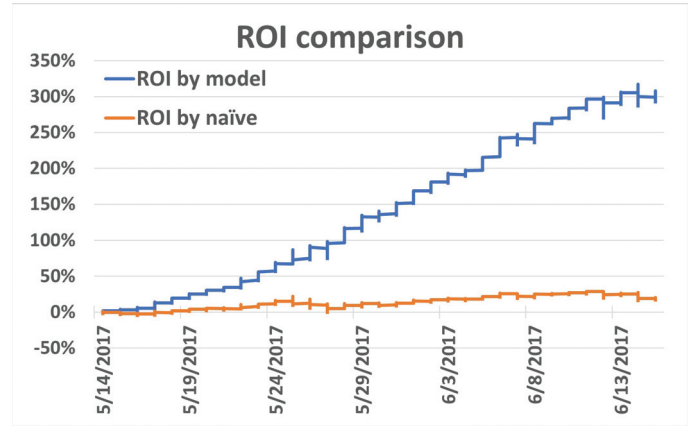


Figure 4. ROI comparison.

6. Conclusions and Implications

Bitcoin market price movement prediction and the dynamic relationship between price and transaction volume distribution are critical for a better understanding of Bitcoin investment, applications, and regulation. In this paper, we present a novel approach to predict Bitcoin market price movement based on the transaction volume distribution. Our results demonstrate that our models significantly improve the accuracy, precision, and recall rates compared to existing models in predicting Bitcoin price movements⁷. Managerially, this is an important finding to inform both investors and regulators of how publicly available transaction data in the blockchain could be utilized to understand the movement of a digital currency. While Bitcoin does benefit from its open decentralized architecture, the public availability of information could provide unfair advantages to some investors.

Author Contributions: Conceptualization, Y.Z., R.G., L.L.G. and P.L.B.; Formal analysis, Y.Z., R.G. and L.L.G.; Investigation, L.L.G.; Methodology, Y.Z., R.G., L.L.G. and P.L.B.; Resources, P.L.B.; Software, Y.Z. and R.G.; Supervision, R.G., L.L.G. and P.L.B.; Writing—original draft, Y.Z., R.G., L.L.G. and P.L.B.; Writing—review & editing, R.G., L.L.G., P.L.B. and A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The University of Texas at Austin Blockchain Initiative and the Center for Risk Management in the McCombs School of Business at the University of Texas at Austin.

Data Availability Statement: The Bitcoin blockchain data are publicly available from multiple sources like Blockchain.com (accessed on 1 March 2020), Blockchair.com (accessed on 1 March 2020), Kaggle.com (accessed on 1 March 2020), and more.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

- Notes
- ¹ In fairness, it should be noted that many “fiat currencies” issued by governments may also sometimes fail to satisfy one or more of these functions.
 - ² However, when there is a lack of trust in the local fiat currency, which sometimes occurs in some developing or emerging economies, or when investment in foreign currency is forbidden, storing value in Bitcoin may be attractive despite its high volatility.
 - ³ Cheah and Fry (2015) showed that Bitcoin exhibits speculative bubbles with a fundamental price of zero .
 - ⁴ We consider T to be the time when a block is mined on the Bitcoin blockchain, which is approximately every 10 min.

- 5 Studies have shown that in many contexts neural network models outperform traditional statistical models for prediction, and that logistic regression is among the best (and easiest to explain) of the of the traditional statistical classification methods (cf., West et al. 1997; Brockett et al. 1994, 2006).
- 6 This is similar to the findings of Chen (2023), who found that older price history was not significant in their predictions, albeit using a different model than ours.
- 7 Our dataset uses Bitcoin information from 2011 to 2017; however, much has evolved in the Bitcoin market since then (e.g., the emergence of Exchange Traded Funds for Bitcoin, El Salvador recognizing Bitcoin as legal tender for transactions, etc.). Nevertheless, there is nothing in these changes that would cast doubt on the conclusion of this paper that there is an accuracy benefit to using segmented transaction data in conjunction with Bitcoin price data to better predict the direction of Bitcoin price movement. Additionally, while we recognize that using data from 2011 to 2017 may raise questions about topicality, we note that using a newer dataset can pose additional challenges because a large number of transactions have been happening off-chain since 2018 (<https://www.chainalysis.com/blog/fake-trade-volume-cryptocurrency-exchanges/> (accessed on 1 March 2020)). Many of the crypto exchanges provided a channel for investors to trade without registering the transaction on the BTC blockchain.

References

- Aalborg, Halvor Aarhus, Peter Molnár, and Jon Erik de Vries. 2019. What can explain the price, volatility and trading volume of bitcoin? *Finance Research Letters* 29: 255–65. [CrossRef]
- Abraham, Jethin, Daniel Higdon, John Nelson, and Juan Ibarra. 2018. Cryptocurrency price prediction using tweet volumes and sentiment analysis. *SMU Data Science Review* 1: No. 3. Available online: <https://scholar.smu.edu/datasciencereview/vol1/iss3/1> (accessed on 10 March 2024).
- Adrian, Tobias, and Tommaso Mancini-Griffoli. 2021. The rise of digital money. *Annual Review of Financial Economics* 13: 57–77. [CrossRef]
- Akcora, Cuneyt G., Asim Kumer Dey, Yulia R. Gel, and Murat Kantarcioglu. 2018. Forecasting bitcoin price with graph chainlets. In *Advances in Knowledge Discovery and Data Mining: 22nd Pacific-Asia Conference, PAKDD 2018, Proceedings, Part III 22, Melbourne, VIC, Australia, June 3–6*. Berlin and Heidelberg: Springer, pp. 765–76.
- Almeida, José, and Tiago Cruz Gonçalves. 2023. Portfolio diversification, hedge and safe-haven properties in cryptocurrency investments and financial economics: A systematic literature review. *Journal of Risk and Financial Management* 16: 3. [CrossRef]
- Alvarez, Fernando, David Argente, and Diana Van Patten. 2023. Are cryptocurrencies currencies? bitcoin as legal tender in el salvador. *Science* 382: eadd2844. [CrossRef]
- Amjad, Muhammad J., and Devavrat Shah. 2016. Nips time series workshop. In *Trading Bitcoin and Online Time Series Prediction, Barcelona, Spain, December 9*. PMLR. pp. 1–15. Available online: <https://proceedings.mlr.press/v55/amjad16.html> (accessed on 1 March 2024).
- Arguelles, Dennys Christian Mallqui. 2018. Predicting the Direction, Maximum, Minimum and Closing Price of Daily /Intra-Daily Bitcoin Exchange Rate Using Batch and Online Machine Learning Techniques. Master's thesis, Universidade Federal de São Carlos, São Carlos, Brazil. Available online: <https://repositorio.ufscar.br/handle/ufscar/10589> (accessed on 1 March 2020)
- Arnosti, Nick, and Matthew Weinberg. 2022. Bitcoin: A natural oligopoly. *Management Science* 68: 4755–71. [CrossRef]
- Atsalakis, George S., Ioanna G. Atsalaki, Fotios Pasiouras, and Constantin Zopounidis. 2019. Bitcoin price forecasting with neuro-fuzzy techniques. *European Journal of Operational Research* 276: 770–80. [CrossRef]
- Bakry, Walid, Audil Rashid, Somar Al-Mohamad, and Nasser El-Kanj. 2021. Bitcoin and portfolio diversification: A portfolio optimization approach. *Journal of Risk and Financial Management* 14: 282. [CrossRef]
- Bariviera, Aurelio F. 2017. The inefficiency of bitcoin revisited: A dynamic approach. *Economics Letters* 161: 1–4. [CrossRef]
- Bartolucci, Silvia, Giuseppe Destefanis, Marco Ortu, Nicola Uras, Michele Marchesi, and Roberto Tonelli. 2020. The butterfly “affect”: impact of development practices on cryptocurrency prices. *EPJ Data Science* 9: 255–65. [CrossRef]
- Bathey, Heather, David Cox, and Michelle Jackson. 2019. On the linear in probability model for binary data. *Royal Society Open Science* 6: 190067. [CrossRef]
- Blau, Benjamin M. 2017. Price dynamics and speculative trading in bitcoin. *Research in International Business and Finance* 41: 493–99. [CrossRef]
- Brockett, Patrick L., Linda L. Golden, Jaeho Jang, and Chuanhou Yang. 2006. A comparison of neural network, statistical methods, and variable choice for life insurers' financial distress prediction. *Journal of Risk and Insurance* 73: 397–419. [CrossRef]
- Brockett, Patrick L., William W. Cooper, Linda L. Golden, and Utai Pitaktong. 1994. A neural network method for obtaining an early warning of insurer insolvency. *The Journal of Risk and Insurance* 61: 402–24. [CrossRef]
- Chan, Louis K. C., and Josef Lakonishok. 1995. The behavior of stock prices around institutional trades. *The Journal of Finance* 50: 1147–74. [CrossRef]
- Cheah, Eng-Tuck, and John Fry. 2015. Speculative bubbles in bitcoin markets? an empirical investigation into the fundamental value of bitcoin. *Economics Letters* 130: 32–36. [CrossRef]
- Chen, Junwei. 2023. Analysis of bitcoin price prediction using machine learning. *Journal of Risk and Financial Management* 16: 51. [CrossRef]
- Dimpfl, Thomas, and Stefania Odelli. 2020. Bitcoin price risk—A durations perspective. *Journal of Risk and Financial Management* 13: 157. [CrossRef]

- Dutta, Aniruddha, Saket Kumar, and Meheli Basu. 2020. A gated recurrent unit approach to bitcoin price prediction. *Journal of Risk and Financial Management* 13: 23. [CrossRef]
- Dyhrberg, Anne H., Sean Foley, and Jiri Svec. 1995. How investible is bitcoin? analyzing the liquidity and transaction costs of bitcoin markets. *Economics Letters* 171: 140–43. [CrossRef]
- Edelen, Roger M., Ozgur S. Ince, and Gregory B. Kadlec. 2016. Institutional investors and stock return anomalies. *Journal of Financial Economics* 119: 472–88. [CrossRef]
- Fischer, Thomas Günter, Christopher Krauss, and Alexander Deinert. 2019. Statistical arbitrage in cryptocurrency markets. *Journal of Risk and Financial Management* 12: 31. [CrossRef]
- Gabaix, Xavier, Parameswaran Gopikrishnan, Vasiliki Plerou, and Harry Eugene Stanley. 2006. Institutional investors and stock market volatility. *Quarterly Journal of Economics* 121: 461–504. [CrossRef]
- Gandal, Neil, J. T. Hamrick, Tyler Moore, and Tali Oberman. 2018. Price manipulation in the bitcoin ecosystem. *Journal of Monetary Economics* 95: 86–96. [CrossRef]
- Geuder, Julian, Harald Kinateder, and Niklas F. Wagner. 2019. Cryptocurrencies as financial bubbles: The case of bitcoin. *Finance Research Letters* 31. . [CrossRef]
- Griffin, John M., and Amin Shams. 2020. Is bitcoin really untethered? *Journal of Finance* 75: 1913–64. [CrossRef]
- Gronwald, Marc. 2019. Is bitcoin a commodity? on price jumps, demand shocks, and certainty of supply. *Journal of International Money and Finance* 97: 86–92. [CrossRef]
- Gsell, Markus. 2008. Assessing the Impact of Algorithmic Trading on Markets: A Simulation Approach (2008). 16th European Conference on Information Systems (ECIS) 2008 Proceedings. 225. Galway, Ireland. Available online: <https://aisel.aisnet.org/ecis2008/225> (accessed on 1 March 2024).
- Holden, Richard, and Anup Malani. 2022. An examination of velocity and initial coin offerings. *Management Science* 68: 9026–41. [CrossRef]
- Jang, Jeewon, and Jangkoo Kang. 2019. Probability of price crashes, rational speculative bubbles, and the cross-section of stock returns. *Journal of Financial Economics* 132: 222–47. [CrossRef]
- Ji, Li Jun, Zhiyong Zhang, and Tiejuan Guo. 2008. To buy or to sell: Cultural differences in stock market decisions based on price trends. *Journal of Behavioral Decision Making* 21: 399–413. [CrossRef]
- Kinderis, Marius, Marija Bezbradica, and Martin Crane. 2018. Bitcoin currency fluctuation. In *Proceedings of the 3rd International Conference on Complexity, Future Information Systems and Risk—Volume 1: COMPLEXIS*. Setúbal: INSTICC, SciTePress, pp. 31–41.
- Kjærland, Frode, Aras Khazal, Erlend A. Krogstad, Frans B. G. Nordstrøm, and Are Oust. 2018. An analysis of bitcoin's price dynamics. *Journal of Risk and Financial Management* 11: 63. [CrossRef]
- Koutmos, Dimitrios. 2018. Bitcoin returns and transaction activity. *Economics Letters* 167: 81–85. [CrossRef]
- Kurbucz, Marcell Tamás. 2019. Predicting the price of bitcoin by the most frequent edges of its transaction network. *Economics Letters* 184: 108655. [CrossRef]
- Kyriazis, Nikolaos A. 2019. A survey on efficiency and profitable trading opportunities in cryptocurrency markets. *Journal of Risk and Financial Management* 12: 67. [CrossRef]
- Li, Jing-Ping, Bushra Naqvi, Syed Kumail Abbas Rizvi, and Hsu-Ling Chang. 2021. Bitcoin: The biggest financial innovation of fourth industrial revolution and a portfolio's efficiency booster. *Technological Forecasting and Social Change* 162: 120383. [CrossRef]
- Lyandres, Evgeny, Berardino Palazzo, and Daniel Rabetti. 2022. Initial coin offering (ico) success and post-ico performance. *Management Science* 68: 8658–79. [CrossRef]
- Madan, Isaac, Shaurya Saluja, and Aojia Zhao. 2014. Automated Bitcoin Trading via Machine Learning Algorithms. Available online: <https://api.semanticscholar.org/CorpusID:14217274> (accessed on 1 March 2020).
- Mai, Feng, Zhe Shan, Qing Bai, Xin (Shane) Wang, and Roger Chiang. 2018. How does social media impact bitcoin value? a test of the silent majority hypothesis. *Journal of Management Information Systems* 35: 19–52. [CrossRef]
- McNally, Sean, Jason Roche, and Simon Caton. 2018. Predicting the price of bitcoin using machine learning. Paper presented at 2018 26th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), Cambridge, UK, March 21–23. pp. 339–43.
- Meese, Richard A., and Kenneth Rogoff. 1983. Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics* 14: 3–24. [CrossRef]
- Miller, Dante, and Jong-Min Kim. 2021. Univariate and multivariate machine learning forecasting models on the price returns of cryptocurrencies. *Journal of Risk and Financial Management* 14: 486. [CrossRef]
- Moosa, Imad, and Kelly Burns. 2014. The unbeatable random walk in exchange rate forecasting: Reality or myth? *Journal of Macroeconomics* 40: 69–81. [CrossRef]
- Nadarajah, Saralees, and Jeffrey Chu. 2017. On the inefficiency of bitcoin. *Economics Letters* 150: 6–9. [CrossRef]
- Niederhoffer, Victor, and Matthew Fontaine Maury Osborne. 1966. Market making and reversal on the stock exchange. *Journal of the American Statistical Association* 61: 897–916. [CrossRef]
- O'Connell, Paul G. J., and Melvyn Teo. 2009. Institutional investors, past performance, and dynamic loss aversion. *Journal of Financial and Quantitative Analysis* 44: 155–88. [CrossRef]
- Popper, Nathaniel. 2018. As Bitcoin Bubble Loses Air, Frauds and Flaws Rise to Surface. *The New York Times*, February 5. Available online: <https://www.nytimes.com/2018/02/05/technology/virtual-currency-regulation.html> (accessed on 1 March 2020).

- Sattarov, Otabek, Heung Seok Jeon, Ryumduck Oh, and Jun Dong Lee. 2020. Forecasting bitcoin price fluctuation by twitter sentiment analysis. Paper presented at 2020 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, November 4–6.
- Scharnowski, Stefan. 2021. Understanding bitcoin liquidity. *Journal of Financial and Quantitative Analysis* 38: 101477. [CrossRef]
- SEC, US Securities Exchange Commission. 2014. Investor Alert: Bitcoin and Other Virtual Currency-Related Investments. Available online: https://www.sec.gov/oiea/investor-alerts-bulletins/investoralertsia_bitcoin (accessed on 1 March 2020).
- Sin, Edwin, and Lipo Wang. 2017. Bitcoin price prediction using ensembles of neural networks. Paper presented at 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Guilin, China, July 29–31. pp. 666–671.
- Su, Chi-Wei, Meng Qin, Ran Tao, and Muhammad Umar. 2020. Financial implications of fourth industrial revolution: Can bitcoin improve prospects of energy investment? *Technological Forecasting and Social Change* 158: 120178. [CrossRef]
- Weaver, Nicholas. 2018. Risks of cryptocurrencies. *Communications of the ACM* 61: 20–24. [CrossRef]
- West, Patricia M., Patrick L. Brockett, and Linda L. Golden. 1997. A comparative analysis of neural networks and statistical methods for predicting consumer choice. *Marketing Science* 16: 370–91. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

On the Realized Risk of Foreign Exchange Rates: A Fractal Perspective

Masoumeh Fathi ¹, Klaus Grobys ^{1,2} and James W. Kolari ^{3,*}

¹ Finance Research Group, School of Accounting and Finance, University of Vaasa, Wolffintie 34, 65200 Vaasa, Finland; masoumeh.fathi@uwasa.fi (M.F.); klaus.grobys@uwasa.fi (K.G.)

² Innovation and Entrepreneurship (InnoLab), University of Vaasa, Wolffintie 34, 65200 Vaasa, Finland

³ Department of Finance, Mays Business School, Texas A&M University, College Station, TX 77843-4218, USA

* Correspondence: jkolari@mays.tamu.edu

Abstract: While well-established literature argues that realized variances are close to a lognormal distribution, this study follows Benoit Mandelbrot by taking a fractal perspective. Using power laws to model realized foreign exchange rate variances, our findings indicate that power laws offer an alternative to the lognormal in terms of goodness-of-fit tests. Further, our analysis shows that estimated power law exponents for seven out of nine realized FX variances are $\hat{\alpha} < 3$, which indicates that the variance of realized variance is statistically undefined. We conclude that the foreign exchange rate market is far riskier than earlier believed. By implication, documented research in an enormous body of literature that draws conclusions from variance analyses stands on shaky grounds.

Keywords: foreign exchange rates; Pareto distributions; power laws; second moment; variance; variance of variance

JEL Classification: C22; G12; G13; G14

Citation: Fathi, Masoumeh, Klaus Grobys, and James W. Kolari. 2024. On the Realized Risk of Foreign Exchange Rates: A Fractal Perspective. *Journal of Risk and Financial Management* 17: 79. <https://doi.org/10.3390/jrfm17020079>

Academic Editor: Thanasis Stengos

Received: 18 January 2024

Revised: 9 February 2024

Accepted: 12 February 2024

Published: 18 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

“There are strong pragmatic reasons to begin the study of economic distributions and time series by those that satisfy the law of Pareto.”

(Benoit Mandelbrot, mathematician known for the *Theory of Roughness and Fractals*)

1. Introduction

Uncertainty in the foreign exchange rate (FX) market has been the subject of intense study in the academic literature. This attention is not surprising in view of the fact that FX market capitalization is considerably larger than that in any other financial market. In 2022, the daily turnover of the global foreign exchange market with 39 different currencies reached 7.5 trillion per day.¹ Previous studies investigating FX risk have employed GARCH-type models (e.g., Jorion 1995; Baillie and Bollerslev 1991; Bollerslev and Melvin 1994).² Despite the widespread usage of GARCH-type models to analyze uncertainty in the FX market, work by Andersen et al. (2003) and Andersen et al. (2004) showed that simple reduced-form time series models for realized volatility (RV) outperform commonly-used GARCH-type models for forecasting future volatility. Unfortunately, many researchers incorporating RV to model FX risk have found there is a problem with realized variances (e.g., Wang and Yang 2009; Corsi 2004; Andersen et al. 2005; Bubák et al. 2011; Corsi et al. 2008). Specifically, Taleb (2020, p. 33) observed that, because the empirical distribution misrepresents the expected realizations in the tails, past data are not useful for predicting future maximums.

Power law functions offer a remedy to the aforementioned problem because the tail exponent of a power law function captures via extrapolation low-probability deviations not seen in the data. Taleb emphasized that such deviations play a disproportionately large share in determining the mean of the distribution. In advocating the usage of power law

functions in empirical finance research, he argued that “There are a lot of theories on why things should be power laws, as sort of exceptions to the way things work probabilistically. But it seems that the opposite idea is never presented: power laws should be the norm, and the Gaussian a special case . . .” Taleb (2020, p. 91).

This study uses power laws to explore the risk of foreign exchange rates in terms of weekly realized variances. Based on a sample of daily prices from 16 May 2006, to 29 December 2023, we compute weekly variances by summing the squared returns over five business days for each currency pair, leaving us with 920 weekly observations. Subsequently, as proposed by Clauset et al. (2009), we employ the maximum likelihood estimation approach to estimate power law exponents. Furthermore, we use the goodness-of-fit (GoF) test derived by these authors for testing various distributions, including lognormal, exponential, and power law. If the p -value for the power law model is higher than for the other competing distributions, the power law model is favored.

Our study contributes to the literature in several ways. First, it extends previous studies on power laws in financial economics. Calvet and Fisher (2004) and Lux et al. (2014) found that power law models usually outperform GARCH models in terms of forecasting future volatility. Some recent studies by Grobys (2021, 2023) used power law functions to model realized variances for foreign exchange rates, the S&P 500, and some commodities. Also, Grobys et al. (2021) employed power law functions to model the realized volatilities of some cryptocurrencies. The commonality among these studies is that the power law null model cannot be rejected for virtually all realized asset market variances. The study by Grobys (2023) found that estimated power law exponents vary between $\hat{\alpha} = 2.2515$ and $\hat{\alpha} = 2.780$, which indicates that the variance of daily realized FX variance does not exist. Further, using an application of Bayes’ rule, he documented that, for the vast majority of realized FX rate variances, the lognormal performs worse in predicting likelihoods for the arrivals of extreme events; hence, the power law model should be favored. The present research adds to this literature by (a) using lower-frequented data; and (b) employing GoF tests to compare various potential distributions. The application of Bayes’ rule in Grobys (2023) only provides evidence on the distribution that is associated with the *most extreme events*. Our approach is complementary in that we explicitly perform statistical tests using probability density functions. Furthermore, Mandelbrot (2008) argued that power law behavior is manifested in invariance of the power law exponent as time frequency changes. Hence, if realized FX variance exhibits fractal-like properties, we would expect that our estimates derived from weekly data should be close to the estimates documented for daily data. Moreover, due to the high rate of false discoveries in financial economics, Hou et al. (2020), Serra-Garcia and Gneezy (2021), and others have recommended scientific replications of reported research. Since we model realized foreign exchange rate variances using a different but related methodology to Grobys (2023), consistent with Hamermesh (2007), we provide replication findings to better understand the fractal-like nature of FX variances.

Relevant to our study, well-established literature has found that the distribution of realized volatility is typically very close to a lognormal distribution (e.g., Andersen et al. 2001a, 2001b; 2003). Renò and Rizza (2003) examined the unconditional volatility distribution of the Italian futures market. They concluded that the standard assumption of lognormal unconditional volatility is rejected and that a much better description is the Pareto distribution. Hence, there is no consensus in the literature on this issue. Is realized variance a lognormal distribution or closer to a power law? To our knowledge, no previous studies explicitly test power laws against other competing distributions to clarify this issue.

Using weekly realized FX variances over the sample from 16 May 2006, to 29 December 2023, our findings indicate that estimated power law exponents vary between $\hat{\alpha} = 2.4834$ and $\hat{\alpha} = 3.5968$. Strikingly, six out of nine realized FX variances have $\hat{\alpha} < 3$, which indicates that the variance of realized variance is statistically undefined. Since realized FX variances show fractal-like behavior, our results are in line with Grobys (2023) and corroborate Mandelbrot’s (2008) arguments. Furthermore, using a significance level of

5%, GoF tests suggest that we cannot reject the power law model for seven out of nine realized FX variances as indicated by p -values $> 5\%$. Interestingly, GoF tests suggest that the lognormal model cannot be rejected for eight out of nine realized FX variances. However, a comparison of the p -values shows that, for five realized FX variances, the p -values for the power law models are substantially higher than the corresponding p -values derived from the lognormal model. Since our findings are consistent with Grobys (2023), we conclude that power laws should be preferred for modeling realized FX risk.

The paper is organized as follows. The next section describes the data. Section 3 discusses the methodology. The last section provides discussion and concluding remarks.

2. Data

The data consist of daily observations downloaded from the global financial website investing.com (accessed on 1 January 2023) for Group of Ten (G10) currency pairs, including AUD/USD, EUR/USD, GBP/USD, NOK/USD, NZD/USD, USD/CAD, USD/CHF, USD/JPY, and USD/SEK. To ensure comparability with earlier documented results, we utilize a sample that is similar to earlier research. Retrieving data from yahoo.com, Grobys (2023) documented that data for the AUD/USD exchange rate are only publicly available from 16 May 2006. Therefore, our study employs data from 16 May 2006, to 29 December 2023. The analysis is restricted to daily data, wherein all FX rates are quoted on the same day, which results in 4599 daily observations. The daily return is computed as follows:

$$RET_{i,t} = 100 * \frac{(P_{i,t+1} - P_{i,t})}{P_{i,t}}, \tag{1}$$

where $RET_{i,t}$ denotes the return of FX i at time t and $P_{i,t}$ denotes the corresponding daily price. The weekly realized variance (WRV) of G10 currency pairs is computed as the sum of the squared returns over five business days for each currency pairs:

$$WRV_{i,j} = \sigma^2_{i,j} = \sum_{t \in j} (RET_{i,t})^2,$$

where j indicates the week and $t \in j$ indicates the corresponding trading days in the respective week. Table 1 reports the descriptive statistics of weekly realized variances for all nine currency pairs. The maximum and minimum values reveal a wide range of realized variances across most of the currency pairs, with USD/CHF showing the largest and EUR/USD the smallest fluctuations. The standard deviations represent the volatility of the realized FX rates variances, with USD/CHF and EUR/USD exhibiting the highest and lowest values, respectively. All currency pairs have high kurtosis due to the presence of fat tails in the distribution of weekly realized variances. Furthermore, among all FX weekly realized variances, skewness is most pronounced for realized USD/CHF variance.

Table 1. Summary statistics for the weekly realized variances of G10 currencies.

	AUD/USD	EUR/USD	GBP/USD	NOK/USD	NZD/USD	USD/CAD	USD/CHF	USD/JPY	USD/SEK
Mean	3.3170	1.6138	1.8063	3.2683	3.2357	1.5839	2.1195	1.9851	2.9710
Median	1.6427	1.0225	1.1017	1.9579	1.8758	0.9303	1.0874	1.0876	1.7309
Std.Dev.	8.7209	1.9674	3.3995	4.8487	4.8140	2.3482	9.2697	3.1163	4.0658
Kurtosis	187.8401	27.0404	289.4048	90.3521	91.2411	64.0870	618.1327	57.6324	25.8182
Skewness	12.3553	4.1499	13.9099	7.5595	7.2678	6.5452	23.5305	5.7512	4.3111
Minimum	0.0676	0.0333	0.0427	0.0398	0.0641	0.0230	0.0081	0.0155	0.0870
Maximum	161.8294	21.9666	78.6839	77.0604	82.8622	31.4039	255.0862	47.5975	39.5884
Obs	920	920	920	920	920	920	920	920	920

This table provides the descriptive statistics of the weekly realized variances for G10 currency pairs, including EUR/USD, AUD/USD, GBP/USD, NOK/USD, NZD/USD, USD/CAD, USD/CHF, USD/JPY, and USD/SEK. The weekly realized variance is com-

puted using daily data from investing.com, spanning from 16 May 2006, to 29 December 2023, and includes 920 observations.

Next, we conduct a visual analysis of the time series plots for each currency pair's realized variance to better understand how they evolved over time. Figure 1 provides a time series plot of the realized FX variances of various currency pairs. The key feature that stands out in these plots is that there are more extreme values or sudden shifts (spikes) than would be expected under normal or lognormal distributions. Some realized FX variances, such as the AUD/USD realized variance and the GBP/USD realized variance, have fewer spikes overall compared to other data series such as the EUR/USD realized variance and USD/JPY realized variance. These results suggest that the former pairs experienced fewer periods of extreme volatility during the sample period. The distribution of spikes over time is not uniform. Some realized FX variances, like the USD/CAD realized variance, show a concentration of extreme events around certain years. For example, there is a noticeable spike around 2008–2009 for the AUD/USD realized variance, which coincides with the global financial crisis. During this period, the Australian dollar had significant volatility arising from its close ties with global commodity prices and risk sentiment. Multiple spikes can be observed for the EUR/USD realized variance from 2010 to 2012, which is likely associated with the European sovereign debt crisis. The sharp spike for the GBP/USD realized variance around 2016 appears to be related to the Brexit referendum in which the United Kingdom voted to leave the European Union. This event caused significant uncertainty and led to a sharp depreciation of the British pound against the US dollar. Additionally, there is some asymmetry in the direction of the spikes. For example, the NOK/USD realized variance and the NZD/USD realized variance show shifts both upward and downward, whereas the EUR/USD realized variance shows predominantly upward shifts. These patterns indicate that certain currencies are more prone to strengthening or weakening against the US dollar under specific conditions. In sum, the presence of fat tails in the graphs indicates that events thought to be very rare may actually occur more frequently than expected. This finding has significant implications for financial risk management, as models based on normal or lognormal distributions may well underestimate the probability of extreme movements and thereby underestimate risk.

Figure 2 shows the difference between the theoretical lognormal distribution (red curve) and the distribution followed by each currency pair's weekly realized variance (black curve). In some cases, especially for currency pairs such as the USD/CHF realized variance and the AUD/USD realized variance, the empirical distribution has a higher peak (kurtosis) and heavier tails as shown by the empirical line extending further along the x-axis than the lognormal line. The presence of heavy tails in the empirical distribution indicates that extreme events—for instance, arising from financial crises, political events, or sudden economic shifts, etc.—leading to large variances in exchange rates are not as rare as the normal or lognormal distribution would suggest. As already mentioned, financial risk management and derivative pricing often wrongly assume lognormality in their models. Additionally, the realized variances of different currency pairs show different levels of variance and different fits to the lognormal distribution. For instance, the EUR/USD realized variance and GBP/USD realized variance have a lower spread, whereas the USD/CHF realized variance shows a much wider spread. These patterns reflect different levels of market volatility and risk associated with each underlying currency pair. Overall, the plots suggest that lognormal distribution is not an adequate model for the distribution of weekly realized variances of these currency pairs.

Figure 3 shows Lognormal quantile–quantile (QQ) plots. They offer a graphical tool to evaluate if a dataset follows lognormal distribution. It is evident that, even though all of the weekly realized variances of the currency pairs show a central tendency that follows a lognormal distribution, they exhibit heavy-tailed behavior as illustrated by the deviations from the theoretical line in the tails. This pattern again is indicative of a fat-tailed distribution, implying a higher frequency of extreme events than what would be expected by a lognormal distribution.

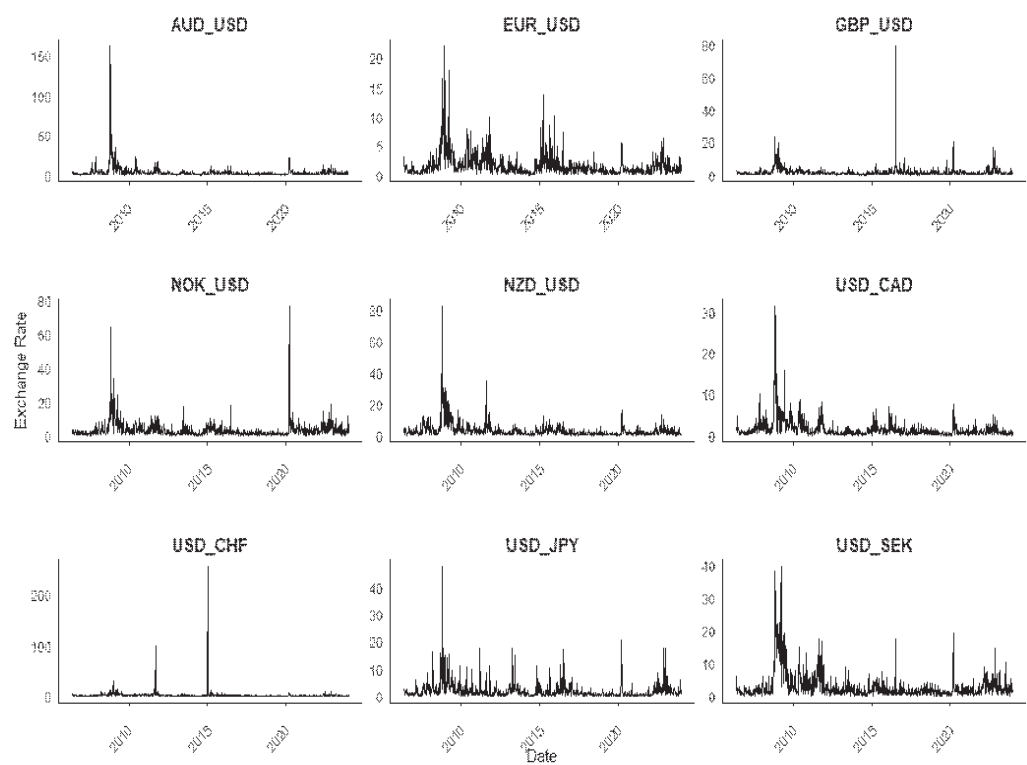


Figure 1. Time series plot.

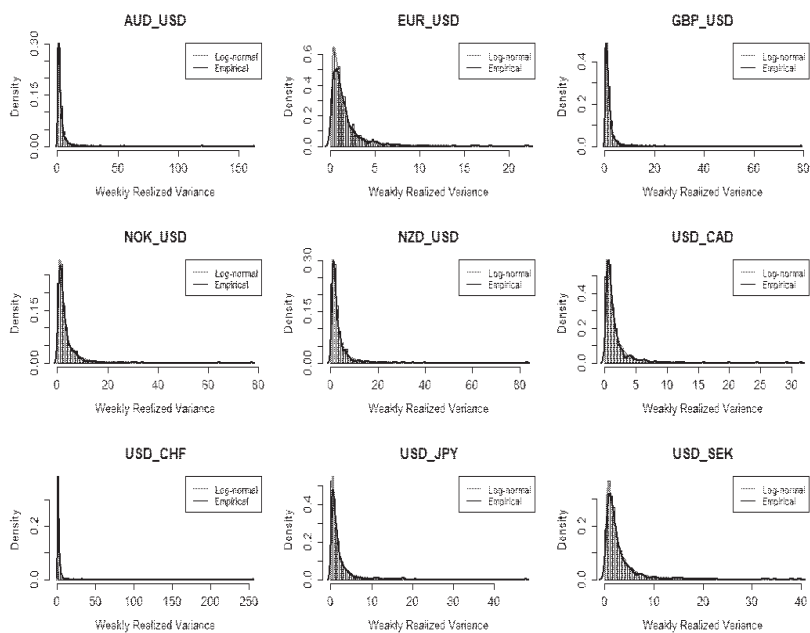


Figure 2. Histogram plot.

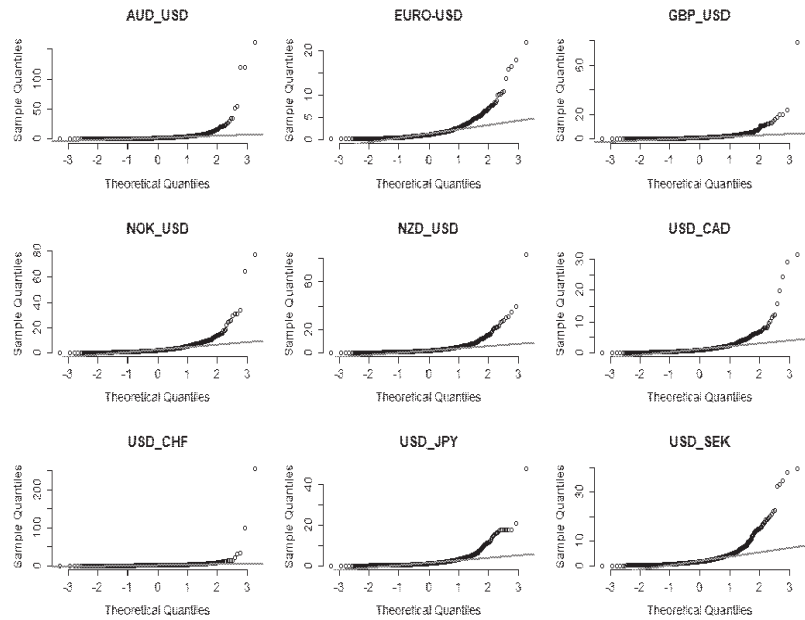


Figure 3. Lognormal QQ plot.

Below are time series graphs of monthly realized variance for weekly realized variances of G10 currency pairs, including EUR/USD, AUD/USD, GBP/USD, NOK/USD, NZD/USD, USD/CAD, USD/CHF, USD/JPY, and USD/SEK, using data from 16 May 2006, to 29 December 2023.

This figure illustrates theoretical lognormal distribution (red curve) versus the empirical distribution (black curve) for the weekly realized variances of G10 currencies, accounting for the following currencies: AUD/USD, EUR/USD, GBP/USD, NOK/USD, NZD/USD, USD/CAD, USD/CHF, USD/JPY, and USD/SEK. The sample is from 16 May 2006, to 29 December 2023.

The figure illustrates the theoretical log-normal quantile (red line) versus the quantile of weekly realized variance for G10 currency pairs including AUD/USD, EUR/USD, GBP/USD, NOK/USD, NZD/USD, USD/CAD, USD/CHF, USD/JPY, and USD/SEK. The sample is from 16 May 2006 to 29 December 2023.

In Figure 4, we employ the mean excess function (ME) plot to evaluate potential power law behaviors for our data. For a given threshold u , the mean excess of a random variable X is defined as:

$$E(X - u | X > u). \quad (2)$$

The presence of a linear upward trend in the ME plot points is consonant with power law behavior, where the slope of the line is directly related to the tail index parameter α . A greater value for α typically denotes a steeper slope on the ME plot. As depicted in Figure 4, the linear trend in the initial part of the plots is suggestive of a distribution with heavy tails, such as a power law distribution.

This figure displays the sorted data on the x-axis against the corresponding values of the mean excess function of weekly realized variances for G10 currency pairs, including AUD/USD, EUR/USD, GBP/USD, NOK/USD, NZD/USD, USD/CAD, USD/CHF, USD/JPY, and USD/SEK, using the whole dataset from 16 May 2006, to 29 December 2023. For a given threshold u of a random variable X , the mean excess is defined as $E(X - u | X > u)$.

Finally, we plot in Figure 5 the cumulative distribution function (CDF) of weekly realized variances for G10 currency pairs. Here, we compare the empirical CDF of each cur-

rency pair’s realized variance with the power law (black), lognormal (red), and exponential (green) distributions. According to Figure 5, the power law and lognormal distributions align closely for AUD/USD and EUR/USD. However, power law provides a superior fit for GBP/USD, NOK/USD, USD/CAD, and USD/CHF, whereas exponential distribution is most suitable for USD/JPY and USD/SEK.

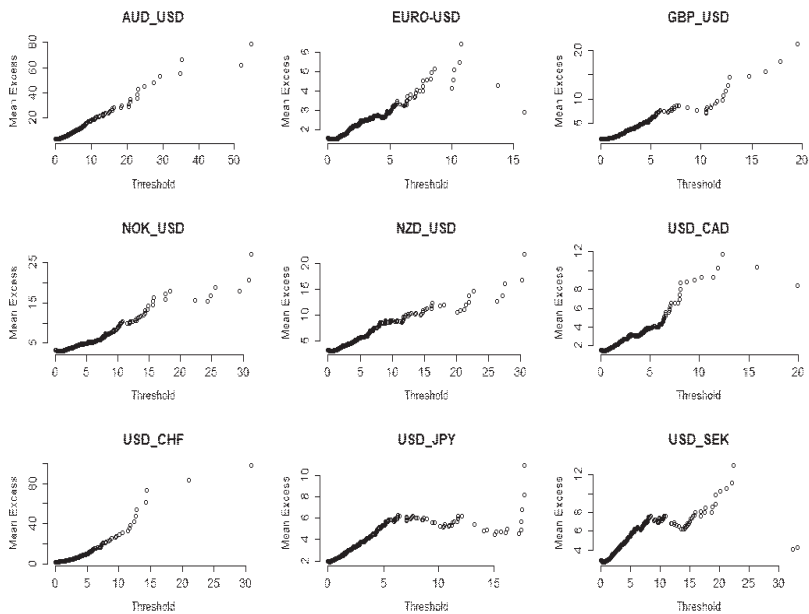


Figure 4. Mean excess plot.

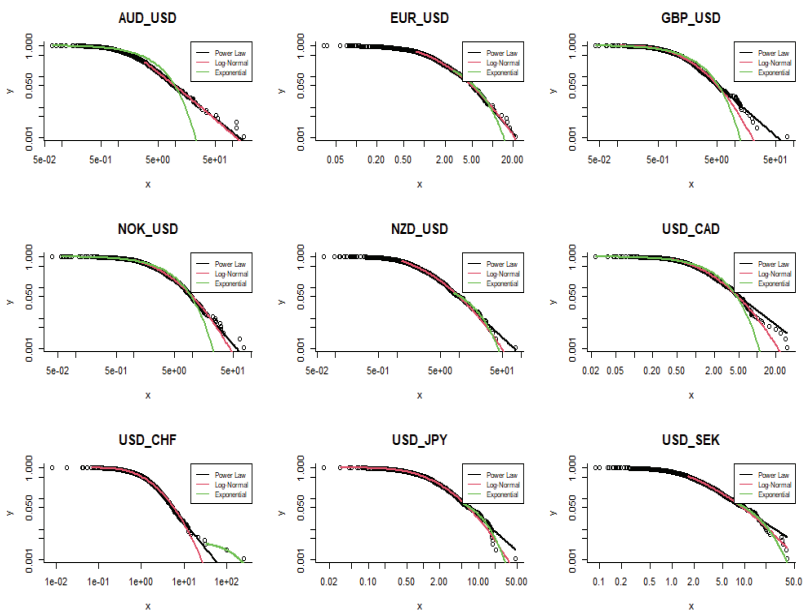


Figure 5. Distribution comparison.

This figure illustrates the cumulative density functions for the weekly realized variances of G10 currency pairs, including AUD/USD, EUR/USD, GBP/USD, NOK/USD, NZD/USD, USD/CAD, USD/CHF, USD/JPY, and USD/SEK, using data from 16 May 2006, to 29 December 2023. The fitted power law (black line), lognormal (red line), and exponential (green line) distributions are also given.

3. Methodology

3.1. Estimating Power Law Exponents Using Maximum Likelihood Estimation

In line with Grobys (2021, 2023), we utilize the power law function to model the variances of foreign exchange rates as:

$$p(x) = Cx^{-\alpha}, \quad (3)$$

where $C = (\alpha - 1)x_{MIN}^{\alpha-1}$ with $\alpha \in \{\mathbb{R}_+ | \alpha > 1\}$, $x = \sigma_i^2$ denotes the respective realized weekly FX variance for currency pair i , provided $x \in \{\mathbb{R}_+ | x_{MIN} \leq x < \infty\}$, x_{MIN} is the minimum value of realized FX variance that is obtained by a power law process, and α is the magnitude of the specific tail exponent.³ The expected value of the realized weekly FX variance, given that $X > x_{MIN}$, is defined as $E[X|X > x_{MIN}]$ and computed by:

$$E[X|X > x_{MIN}] = \int_{x_{MIN}}^{\infty} xp(x)dx = \frac{(\alpha - 1)}{(\alpha - 2)}x_{MIN}. \quad (4)$$

The second moment, which is the variance of the realized variance, given that $X > x_{MIN}$, is defined as $E[X^2|X > x_{MIN}]$ and computed by:

$$E[X^2|X > x_{MIN}] = \int_{x_{MIN}}^{\infty} x^2p(x)dx = \frac{(\alpha - 1)}{(\alpha - 3)}x_{MIN}^2. \quad (5)$$

Finally, higher moments of order ‘k’ are computed as:

$$E[X^k|X > x_{MIN}] = \frac{(\alpha - 1)}{(\alpha - 1 - k)}x_{MIN}^k \quad (6)$$

Based on Equations (4) and (5), it is evident that the theoretical mean of realized variances is only defined for $\alpha > 2$, whereas the variance of the realized variance requires $\alpha > 3$.

Following White et al. (2008) and Clauset et al. (2009), who argued that maximum likelihood estimation (MLE) is the optimal method for calculating power law exponents, the tail exponents $\hat{\alpha}$ are computed as follows:

$$\hat{\alpha} = 1 + N \left(\sum_{i=1}^N \ln \left(\frac{x_i}{x_{MIN}} \right) \right)^{-1}, \quad (7)$$

where $\hat{\alpha}$ denotes the MLE estimator, $N \leq T$ is the number of observations greater than x_{MIN} , and the rest of notation is as earlier. According to Equation (7), the MLE estimator depends on the chosen x_{MIN} . Like Clauset et al. (2009), we select x_{MIN} by minimizing the distance between the power law model and the empirical data defined as Kolmogorov–Smirnov distance (D) or maximum distance between the cumulative density functions (CDFs) of the data and the fitted model:

$$D = \text{MAX}_{x \geq x_{MIN}} |S(x) - P(x)| \quad (8)$$

where $S(x)$ represents the cumulative distribution function (CDF) of the data for observations with values greater than or equal x_{MIN} , and $P(x)$ is the CDF for the power law model that provides the best fit to the data in the range of $x \geq x_{MIN}$. The estimate $\hat{x}_{MIN}$ is

determined as the value of x_{MIN} that minimizes D . Moreover, the standard deviation for estimated power law exponents is defined as:

$$\sigma = \frac{\hat{\alpha} - 1}{\sqrt{N}} + O\left(\frac{1}{N}\right). \tag{9}$$

Summary statistics for the estimated power law exponents are presented in Table 2. The estimated tail exponent for the realized AUD/USD variance is $\hat{\alpha} = 2.4834$, with corresponding standard deviation $\hat{\sigma} = 0.0911$. It follows that $\hat{\alpha}$ is 5.67 standard deviations below a hypothesized tail exponent corresponding to $\alpha = 3$. Therefore, α for the realized AUD/USD variance is statistically significant $\alpha < 3$ indicating that the variance of realized variance is statistically undefined. The same argument holds for the realized variances of USD/CAD, USD/GBP, USD/CHF, USD/JPY, and USD/SEK; however, the estimated tail exponent for the realized NZD/USD variance (e.g., $\hat{\alpha} = 2.8648$) is statistically not different from $\alpha = 3$. Furthermore, we observe from Table 2 for the lower bounds of the 95% confidence intervals, $\alpha < 3$ holds for all realized variance series, which provides further evidence that the infinite variance hypothesis cannot be rejected. Overall, we infer from these findings that the theoretical variances for the distributions of at least six out of nine realized FX variances are undefined, consistent with Grobys’ (2023) findings derived from daily data (viz., estimated power law exponents for daily range-based FX variances varied between $\hat{\alpha} = 2.2515$ and $\hat{\alpha} = 2.7807$).

Table 2. Estimated power law functions.

Distribution	$\hat{\alpha}$	$\hat{\sigma}$	x_{MIN}	D	N_{PL}	95% CI
AUD/USD	2.4834	0.0911	2.9452	0.0242	28.80%	[2.3048, 2.6620]
EUR/USD	3.5968	0.3126	4.3242	0.0497	7.50%	[2.9841, 4.2095]
GBP/USD	2.6704	0.1092	1.9530	0.0319	25.43%	[2.4563, 2.8844]
NOK/USD	3.1845	0.1978	6.1729	0.0389	13.26%	[2.7968, 3.5721]
NZD/USD	2.8648	0.1617	5.4697	0.0509	14.46%	[2.5479, 3.1818]
USD/CAD	2.5048	0.0869	1.4408	0.0535	32.61%	[2.3345, 2.6751]
USD/CHF	2.7312	0.1256	2.4059	0.0275	20.65%	[2.4851, 2.9774]
USD/JPY	2.5888	0.1276	3.0011	0.0485	16.85%	[2.3387, 2.8390]
USD/SEK	2.5553	0.1019	3.2868	0.0462	25.33%	[2.3556, 2.7550]

This table reports the results of estimating power law exponents for weekly realized variances for G10 currencies, including AUD/USD, EUR/USD, GBP/USD, NOK/USD, NZD/USD, USD/CAD, USD/CHF, USD/JPY, and USD/SEK, using data from 16 May 2006, to 29 December 2023. The estimate $\hat{\alpha}$ represents the estimated tail exponent and $\hat{\sigma}$ is the estimated standard deviation. The lower threshold x_{MIN} is estimated using the optimized Kolmogorov–Smirnov distance (D) distance as proposed by Clauset et al. (2009). The estimated x_{MIN} is the value that corresponds to optimal distance D . N_{PL} denotes the percentage of sample observations governed by a power law process. The last column reports the 95% confidence interval (CI) intervals for the estimated power law exponents.

3.2. Testing the Power Law Model

According to Clauset et al. (2009), a tail exponent within the interval $2 < \alpha < 3$ does not necessarily indicate that the power law is a plausible fit to the data. Therefore, we need to verify whether the observed dataset genuinely adheres to a power law distribution. This test employs the Kolmogorov–Smirnov distance (D) as outlined in Equation (8). The objective is to determine whether the empirical data and synthetic data generated from the power law distribution, with specific x_{MIN} and α values, originate from the same distribution. The null hypothesis for the goodness-of-fit test hypothesizes that the empirical data and the synthetic data generated from the power law distribution, with specified x_{MIN} and α values, share the same distribution. If so, power law distribution is a plausible fit for the empirical data. Conversely, the alternative hypothesis posits that the

empirical data and the synthetic data do not originate from the same distribution, which indicates that power law distribution is not a suitable fit for the empirical data. In this case, alternative distribution models should be explored to better characterize the underlying data distribution.

To implement this analysis, we generate large synthetic datasets followed by the power law distribution utilizing the estimated parameters x_{MIN} and α derived from the previous section for each FX currency variances. Specifically, we compute the p -value of this test by fraction $\#(D_S > D)/K$, where D_S represents the estimated Kolmogorov–Smirnov distance for a synthetic dataset with specified x_{MIN} and α values mirroring those estimated for the original FX variances, D denotes the estimated Kolmogorov–Smirnov distance for the original dataset, and K is the number of synthetic datasets generated for testing purposes.

We set the significance level at 5%, which means that we do not reject the power law model if $\# \frac{D_S > D}{K} > 0.05$. This test aims to determine whether the empirical data and the generated dataset share the same distribution. We use $K = 500$ synthetic data series to compute the p -values for the goodness-of-fit tests. The second column of Table 3 reports the p -values corresponding to the null hypothesis of the power law model. It is evident that, for all realized FX variances, the p -values are above 5%, with the exception of USD/CAD and USD/SEK. These findings strongly support the notion that power laws are indeed plausible distributions for modeling realized FX variances.

Table 3. Goodness-of-fit test.

	Power Law	Lognormal	Exponential
EUR/USD	0.96	0.05	0.00
AUD/USD	0.76	0.40	0.08
GBP/USD	0.64	0.38	0.14
NOK/USD	0.86	0.60	0.00
NZD/USD	0.30	0.90	0.10
USD/CAD	0.00	0.46	0.00
USD/CHF	0.98	0.26	0.44
USD/JPY	0.16	0.64	0.32
USD/SEK	0.04	0.86	0.40

This table reports the p -values from goodness-of-fit tests derived from the Kolmogorov–Smirnov distance to examine if our empirical data and the generated data from the power law distribution with particular x_{min} and α belong to the same distribution. The corresponding p -values are reported in the second column. The table gives the p -values for goodness-of-fit tests for exponential and lognormal distributions in the third and fourth columns, respectively. We use $K = 500$ synthetic data series to compute the p -values.

3.3. An Empirical Comparison of Distributions

Although our analysis in the previous section suggests that power law distributions provide a reasonable fit for realized FX variances, it should be noted that they may not be the optimal model, and other distributions could potentially offer a superior fit. Consequently, we focus on comparing two primary candidates—that is, the lognormal and exponential distributions against the power law to determine the appropriate fitting model for our data. It is important to note that prior studies have documented that realized volatility tends to follow a distribution that is nearly lognormal (e.g., Andersen et al. 2001a; 2001b, 2003). In our study, we investigate the suitability of the lognormal and exponential distributions for realized variance of G10 currency pairs. Employing a goodness-of-fit test similar to the one described by Clauset et al. in the previous section, we test the hypothesis that the lognormal (exponential) distribution can be considered a plausible fit. Thus, the lognormal (exponential) distribution is used as the null hypothesis against which the currency pair variances are tested at the 5 percent significance level as before. Columns 3 and 4 in Table 3 display the p -values of the goodness-of-fit test for lognormal and exponential distributions,

respectively. The results in column 3 indicate that eight out of nine p -values exceed the significance level of 5%, implying that lognormal distribution is a plausible fit to the data, with the exception of the EUR/USD realized variance. In addition, the results in column 4 indicate that exponential distribution is at least a plausible fit for the realized variances of the following currency pairs: AUD/USD, GBP/USD, NZD/USD, USD/CHF, USD/JPY, and USD/SEK.

According to Clauset et al., using p -values for power law models alongside other potential alternative distributions (e.g., lognormal and exponential) allows us to construct a persuasive argument regarding the suitability of the power law model for all currency pairs. Specifically, they observed that a high p -value for the power law model in combination with lower p -values for other distributions supports ruling out alternative distributions. We find in Table 3 that power law models for the realized variances of EUR/USD, AUD/USD, GBP/USD, NOK/USD, and USD/CHF exhibit considerably higher p -values than the alternative distributions. Interestingly, Grobys (2023) employed an application of Bayes' rule to examine the conditional probability that the underlying distribution is lognormally distributed, given that distribution-specific extreme events occurred. In doing so, he explored the cross-section of extreme events for daily realized range-based FX variances. His findings indicated that, when lognormal distribution is assumed and power law processes are equally likely, for six out of nine daily realized FX rate variances, the probability that the underlying distribution is lognormal given the arrival of the corresponding maximums is <25%. Even when it was assumed that lognormal distribution is 70% likely, for six out of nine daily realized FX rate variances, the probability that the underlying distribution is lognormal given the arrival of the corresponding maximums is <40%. Overall, Grobys (2023) concluded that, for the vast majority of realized FX rate variances, we can rule out a lognormal distribution as the corresponding underlying data-generating process for realized FX variances. The results of the present study in association with the results documented in Grobys (2023) provide evidence that power law models describe the distribution of realized FX variances more accurately. This inference is in line with Renò and Rizza (2003) but contradicts some well-established literature (Andersen et al. 2001a; 2001b; 2003).

3.4. Robustness Checks

Table A1 in the Appendix A reports the results from a sample-split analysis of estimating power law exponents. In this analysis, we divided the sample into two non-overlapping subsamples of equal length. Panel A of Table A1 reports the estimation results for the first subsample from 16 May 2006, to 3 March 2015, whereas Panel B documents the estimation results for the second subsample from 10 March 2015, to 29 December 2023. Each subsample includes 460 observations. From Table A1, it is evident that the point estimates $\hat{\alpha}_{1i}$ and $\hat{\alpha}_{2i}$ are fairly similar between the two subsamples. However, it is noteworthy that there are some differences between the two subsamples. For instance, subsample 1—covering the period up to 3 March 2015—includes the aftermath of the 2008 financial crisis and the European debt crisis, among other events. These severe economic events are likely to have influenced market volatility and, in turn, the behavior of currency variances. On the other hand, subsample 2 captures a different economic period, including the gradual recovery from previous crises, shifts in monetary policy across major economies, and the impact of the COVID-19 pandemic. Overall, from Panels A and B of Table A1 for the lower bounds of the 95% confidence intervals, it holds that $\alpha < 3$ for all realized variance series, which is similar to our previously documented evidence.

4. Discussion and Concluding Remarks

The present study applied a fractal perspective to investigate the second moment of foreign exchange rates. We examined G10 currencies that account for approximately 70% of overall FX market capitalization. Instead of relying on parametric GARCH-type models, we utilized realized variances derived from daily price data. Realized variances are employed in MLE estimations, which provide reliable estimation even in the presence

of extremely fat-tailed data. Well-established literature on realized volatility contends that the underlying distribution of realized asset uncertainty is close to normal (e.g., Andersen et al. 2001a; 2001b; 2003). Our findings contradict this view as power law models provide better data fits and more reliable probabilities for the arrivals of extreme events. We infer that the conventional lognormal distribution for modeling risk may severely underestimate tail risks.

Further findings indicated that the variance of variance does not exist for most of the foreign exchange rate variances under study. This evidence is in line with Grobys (2023), who fitted power laws to daily range-based FX variances. Mandelbrot (2008) argued that power law behavior is manifested in invariance of the power law exponent across time frequencies. In his seminal study on cotton price changes, Mandelbrot (1963) was the first to show that the power law exponent for cotton price changes does not change across various time frequency—a property suggesting fractal-like behavior. Consistent with his findings, weekly data on realized FX variances behave very similar to daily data.

An important implication of our empirical results is that standard statistical methodologies can lead to invalid inferences. In this regard, Fama (1963) observed that OLS regression assuming finite variances may not be appropriate in empirical analyses in financial economics due to infinite variances. Using a novel methodology that explicitly focuses on the statistical information residing in the tails, our findings suggested that the variances of the majority of exchange rate variance are infinite. As noted by Mandelbrot (2008), the reliance on traditional finance models gives a comforting impression of precision and competence:

“It is false confidence, of course. The problem lies at the roots of the standard model, in its assumption that the best way to think about stock markets is as a grand game of coin-tossing. If you are going to use probability to model financial markets, then you had better use the right kind of probability. Real markets are wild. Their price fluctuations can be hair-raising—far greater and more damaging than the mild variations of orthodox finance.” (Mandelbrot 2008, p. 105).

This study supports Mandelbrot’s (2008) viewpoint that price fluctuations in the currency market are wild—indeed so uncertain that the second moment of realized variances does not converge.

While we focused on G10 currencies in the FX market, future research is recommended to explore this issue in other financial markets, such as the market for cryptocurrencies which is exposed to greater risks than established financial markets. Also, it is possible that, as documented in Ibragimov et al. (2013), power law behavior could be more pronounced in emerging than advanced economies. Hence, future research is warranted to investigate this issue for emerging economies.

Finally, Gabaix and Ibragimov (2011) proposed various approaches for estimating the optimal tail exponent. Several studies have concluded that inferences based on the tail index using a maximum likelihood estimator suffer from shortcomings, including sensitivity to dependence and small sample sizes (e.g., Embrechts et al. 1997). Even though our sample consists of sufficient numbers of realized variance observations, dependency could exist in our data. We chose the MLE approach in line with Clauset et al. (2009) to (1) ensure that our results are comparable to previous studies and (2) implement goodness-of-fit tests to discriminate between competing distributions. However, following Gabaix and Ibragimov, future studies are encouraged to use the log–log rank size OLS regression approach rather than the MLE approach.

Author Contributions: Conceptualization: M.F., K.G. and J.W.K. Methodology: M.F. and K.G. Software: M.F. Validation: M.F., K.G. and J.W.K. Formal analysis: M.F. Investigation: M.F., K.G. and J.W.K. Resources: M.F., K.G. and J.W.K. Data curation: M.F. Writing—original draft preparation: M.F., K.G. and J.W.K. Writing—review and editing: J.W.K. Visualization: M.F. and J.W.K. Supervision: K.G. and J.W.K. Project administration: M.F., K.G. and J.W.K. Funding acquisition: N/A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data used in this study was downloaded for free from the data base investing.com.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Sample-split analysis.

Panel A						
Distribution	$\hat{\alpha}$	$\hat{\sigma}$	x_{MIN}	D	N_{PL}	95% CI
AUD/USD	2.4139	0.1666	6.2002	0.0545	15.65%	[2.0873, 2.7405]
EUR/USD	2.4325	0.0934	1.3364	0.0691	51.09%	[2.2493, 2.6156]
GBP/USD	2.5637	0.1196	1.4831	0.0416	37.17%	[2.3293, 2.7981]
NOK/USD	3.2158	0.2432	6.0219	0.0553	18.04%	[2.7391, 3.6926]
NZD/USD	2.4391	0.1085	3.4412	0.0582	38.26%	[2.2265, 2.6518]
USD/CAD	3.0643	0.2522	3.8276	0.0632	14.57%	[2.5700, 3.5586]
USD/CHF	2.7203	0.1640	2.9686	0.0462	23.91%	[2.3989, 3.0418]
USD/JPY	2.4901	0.1268	2.3796	0.0452	30.00%	[2.2415, 2.7387]
USD/SEK	2.3217	0.1058	3.2904	0.0736	33.91%	[2.1143, 2.5291]
Panel B						
Distribution	$\hat{\alpha}$	$\hat{\sigma}$	x_{MIN}	D	N_{PL}	95% CI
AUD/USD	3.3773	0.2389	3.0557	0.0453	21.52%	[2.9090, 3.8456]
EUR/USD	2.7133	0.1342	1.1669	0.0481	35.43%	[2.4503, 2.9764]
GBP/USD	2.7723	0.1682	2.0690	0.0476	24.13%	[2.4426, 3.1020]
NOK/USD	2.8486	0.1591	3.1635	0.0389	29.34%	[2.5367, 3.1604]
NZD/USD	3.1287	0.2139	2.9721	0.0732	21.52%	[2.7094, 3.5481]
USD/CAD	3.0508	0.1772	1.2762	0.0650	29.13%	[2.7036, 3.3981]
USD/CHF	3.0085	0.1804	1.3947	0.0474	26.95%	[2.6550, 3.3620]
USD/JPY	2.8305	0.2538	3.2205	0.0528	11.30%	[2.3330, 3.3281]
USD/SEK	3.4147	0.2491	3.0119	0.0689	20.43%	[2.9266, 3.9029]

This table presents the results from a sample-split analysis for estimating power law exponents. We divide the sample period into two equal-length subsamples. Panel A details the first subsample from 16 May 2006, to 3 March 2015, whereas Panel B covers the second subsample from 10 March 2015, to 29 December 2023. Each subsample includes 460 observations. $\hat{\alpha}$ represents the estimated tail exponent, and $\hat{\sigma}$ is the estimated standard deviation. The lower threshold x_{MIN} is estimated using the optimized Kolmogorov–Smirnov distance (D) distance as proposed by Clauset et al. (2009). The estimated x_{MIN} is the value that corresponds to optimal distance D . N_{PL} denotes the percentage of sample observations governed by a power law process. The last column reports the 95% confidence interval (CI) intervals for the estimated power law exponents.

Notes

- 1 See <https://www.statista.com/statistics/247328/activity-per-trading-day-on-the-global-currency-market/> (accessed on 1 January 2023).
- 2 Other related studies are Alexander (1995) and Bauwens et al. (2005), who used (G)ARCH-type models to explore common volatility in the foreign exchange rate market as well as the impact of scheduled and unscheduled news announcements on foreign exchange rate return volatility.
- 3 Note that for the sake of readability, we skip in our notation here the index j denoting the week.

References

Alexander, Carol O. 1995. Common volatility in the foreign exchange market. *Applied Financial Economics* 5: 1–10. [CrossRef]

Andersen, Torben G., Tim Bollerslev, and Francis X. Diebold. 2005. *Roughing It Up: Including Jump Components in the Measurement, Modelling and Forecasting of Return Volatility*. NBER Working Paper No. 11775. Cambridge: National Bureau of Economic Research.

Andersen, Torben G., Tim Bollerslev, and Nour Meddahi. 2004. Analytical evaluation of volatility forecasts. *International Economic Review* 45: 1079–110. [CrossRef]

- Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Heiko Ebens. 2001a. The distribution of realized stock return volatility. *Journal of Financial Economics* 61: 43–76. [CrossRef]
- Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys. 2001b. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* 96: 42–55. [CrossRef]
- Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys. 2003. Modeling and forecasting realized volatility. *Econometrica* 71: 579–625. [CrossRef]
- Baillie, Richard T., and Tim Bollerslev. 1991. Intra-day and inter-market volatility in foreign exchange rates. *Review of Economic Studies* 58: 565–85. [CrossRef]
- Bauwens, Luc, Walid Ben Omrane, and Pierre Giot. 2005. News announcements, market activity and volatility in the euro/dollar foreign exchange market. *Journal of International Money and Finance* 24: 1108–25. [CrossRef]
- Bollerslev, Tim, and Michael Melvin. 1994. Bid—Ask spreads and volatility in the foreign exchange market: An empirical analysis. *Journal of International Economics* 36: 355–72. [CrossRef]
- Bubák, Vít, Evžen Kočenda, and Filip Žikeš. 2011. Volatility transmission in emerging European foreign exchange markets. *Journal of Banking and Finance* 35: 2829–41. [CrossRef]
- Calvet, Laurent E., and Adlai J. Fisher. 2004. Regime-switching and the estimation of multifractal processes. *Journal of Financial Econometrics* 2: 44–83. [CrossRef]
- Clauset, Aaron, Cosma Rohilla Shalizi, and Mark E. J. Newman. 2009. Power law distributions in empirical data. *SIAM Review* 51: 661–703. [CrossRef]
- Corsi, Fulvio. 2004. *A Simple Long Memory Model of Realized Volatility*. Working Paper. Lugano: Institute of Finance, University of Lugano.
- Corsi, Fulvio, Stefan Mittnik, Christian Pigorsch, and Uta Pigorsch. 2008. The volatility of realized volatility. *Econometric Reviews* 27: 46–78. [CrossRef]
- Embrechts, Paul, Claudia Klüppelberg, and Thomas Mikosch. 1997. *Modelling Extremal Events for Insurance and Finance*. New York: Springer.
- Fama, Eugene F. 1963. Mandelbrot and the stable Paretian hypothesis. *Journal of Business* 36: 420–29. [CrossRef]
- Gabaix, Xavier, and Rustam Ibragimov. 2011. Rank—1/2: A simple way to improve the OLS estimation of tail exponents. *Journal of Business and Economic Statistics* 29: 24–39. [CrossRef]
- Grobys, Klaus. 2021. What do we know about the second moment of financial markets? *International Review of Financial Analysis* 78: 101891. [CrossRef]
- Grobys, Klaus. 2023. Correlation versus co-fractality: Evidence from foreign-exchange-rate variances. *International Review of Financial Analysis* 86: 102531. [CrossRef]
- Grobys, Klaus, James W. Kolari, and Niranjana Sapkota. 2021. On the stability of stablecoins. *Journal of Empirical Finance* 64: 207–23. [CrossRef]
- Hamermesh, Daniel S. 2007. Viewpoint: Replication in economics. *Canadian Journal of Economics* 40: 715–33. [CrossRef]
- Hou, Kewei, Chen Xue, and Lu Zhang. 2020. Replicating anomalies. *Review of Financial Studies* 33: 2019–133. [CrossRef]
- Ibragimov, Marat, Rustam Ibragimov, and Paul Kattuman. 2013. Emerging markets and heavy tails. *Journal of Banking and Finance* 37: 2546–59. [CrossRef]
- Jorion, Philippe. 1995. Predicting volatility in the foreign exchange market. *Journal of Finance* 50: 507–28. [CrossRef]
- Lux, Thomas, Leonardo Morales-Arias, and Cristina Sattarhoff. 2014. A Markov-switching multifractal approach to forecasting realized volatility. *Journal of Forecasting* 33: 532–41. [CrossRef]
- Mandelbrot, Benoit B. 1963. The variation of certain speculative prices. *Journal of Business* 36: 394–419. [CrossRef]
- Mandelbrot, Benoit B. 2008. *The (Mis)behavior of Markets. A Fractal View of Risk, Ruin and Reward*. London: Profile Books.
- Renò, Roberto, and Rosario Rizza. 2003. Is volatility lognormal? Evidence from Italian futures. *Physica A: Statistical Mechanics and Its Applications* 322: 620–28. [CrossRef]
- Serra-Garcia, Marta, and Uri Gneezy. 2021. Nonreplicable publications are cited more than replicable ones. *Science Advances* 7: 1705. [CrossRef]
- Taleb, Nassim Nicholas. 2020. *Statistical Consequences of Fat Tails: Real World Preasymptotics, Epistemology, and Applications*. Papers and Commentary, STEM. Cambridge: Academic Press.
- Wang, Jianxin, and Minxian Yang. 2009. Asymmetric volatility in the foreign exchange markets. *Journal of International Financial Markets, Institutions and Money* 19: 597–615. [CrossRef]
- White, Ethan P., Brian J. Enquist, and Jessica L. Green. 2008. On estimating the exponent of power law frequency distributions. *Ecology* 89: 905–12. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

Volatility and Herding Bias on ESG Leaders' Portfolios Performance

Nektarios Gavrilakis and Christos Floros *

Department of Accounting and Finance, Hellenic Mediterranean University, 71410 Heraklion, Crete, Greece; ngavrilakis@hmu.gr

* Correspondence: cfloros@hmu.gr

Abstract: We here analyze the factor loadings given by the CAPM, the Fama–French three (FF3), and the five-factor model (FF5), and test the performance and the validity of adding two more factors (volatility and dispersion of returns) to the FF5 factor model of European index-based ESG leaders' portfolios. Our ESG leaders' portfolios generated significant negative alphas during 2012–2022, corroborating the literature's negative argument. The negative abnormal returns of ESG leaders' portfolios are homogeneous across the three ESG pillars. We conclude that European ESG leaders' portfolios are biased toward large cap and value stocks with robust operating profitability and against aggressive investments. As robustness tests, we examine Global ESG leaders' index-based portfolios, producing the same results but with reduced importance in some loading factors like profitability and investment strategy. Furthermore, we deduced that European and Global ESG leaders' portfolios tilt towards volatility and herding bias.

Keywords: ESG; Jensen's alpha; volatility; herding

1. Introduction

The Environmental Social and Governance (ESG) market is expected to have increased its assets under management (AuM) to USD 33.9 tn by 2026, from USD 18.4 tn in 2021 (Ross 2023). PricewaterhouseCoopers (2022) assumes ESG investing will be a potential opportunity in the coming years. European investors develop increased demand for ESG products despite headwinds due to rising interest rates, constant high inflation, and recession fears (Morningstar Manager Research 2023). To advance the effort toward ESG investing in the sustainable financial sector, it is essential to understand the effects of ESG leaders on portfolio performance.

Our study contributes to the relatively limited research on sustainable investing, and addresses whether investing in European ESG leaders could yield positive risk-adjusted returns. Beyond measuring Jensen's alpha against the market benchmark and the factor loadings given by the Capital Asset Pricing Model (Treynor 1961, 1962), the Fama and French (1993) three-factor model (FF3) and the Fama and French (2015) five-factor model (FF5), we analyze the performance and the validity of adding two more factors (volatility and dispersion of returns) to the FF5 model. In this way, we might address two more issues: if ESG leaders' portfolios are biased toward volatility and dispersion of returns and if ESG leaders' portfolios lean toward herding.

As proxies for ESG leaders, firstly, we selected a sample consisting of European ESG leaders' indices (Stoxx Europe ESG Environmental Leaders, Stoxx Europe ESG Social Leaders, Stoxx Europe ESG Governance Leaders, Stoxx Europe ESG Leaders select 30), and as a benchmark, we selected the Stoxx Europe 600 Index. Secondly, for the robustness check, we selected a sample consisting of Global ESG leaders' indices (Stoxx Global ESG Environmental Leaders, Stoxx Global ESG Social Leaders, Stoxx Global ESG Governance Leaders, Stoxx Global ESG Leaders), and as a benchmark, we selected the Stoxx Global 1800 index. Data were obtained from the Qontigo (2023) database, which uses ESG scores

Citation: Gavrilakis, Nektarios, and Christos Floros. 2024. Volatility and Herding Bias on ESG Leaders' Portfolios Performance. *Journal of Risk and Financial Management* 17: 77. <https://doi.org/10.3390/jrfm17020077>

Academic Editor: Thanasis Stengos

Received: 15 January 2024

Revised: 9 February 2024

Accepted: 12 February 2024

Published: 16 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

in environmental, social, and governance pillars from the Morningstar/Sustainalytics (2023) provider “<https://www.sustainalytics.com/esg-data> (accessed on 7 July 2023)”. The limited availability of data on EGS index daily prices forced us to use the period from 1 June 2012 to 14 July 2022 from the above provider. de Oliveira et al. (2020) argued that ESG indices are a valuable investors’ tool that can be used to reduce the fear of increased uncertainty.

The literature on ESG and portfolio performance is developing, motivating us to focus on this area. Some studies proposed ESG integration as a tool for portfolio optimization (Henriksson et al. 2019) and confirm superior returns (Renneboog et al. 2008; Consolandi et al. 2009), while others (Bauer et al. 2005; Adler and Kritzman 2008; and Berlinger and Lovas 2015) show evidence of underperformance in ESG portfolios. Studies by Managi et al. (2012) and Hartzmark and Sussman (2019) support neutrality, meaning that ESG markets are efficient, and managing positive risk-adjusted returns is impossible. To help address the effect of ESG tilting on ESG leaders’ performance, we examine the factor loadings given by the CAPM, the FF3, and the FF5 models, and provide some practical implications for corporate decision-making and portfolio construction processes.

Our second motivation is to identify the validity of adding two more factors (volatility and dispersion of returns) to the FF5 factor model. Volatility behavior is a dominant factor when selecting portfolio construction and performance assets. Like the Chicago Board Options Exchange (CBOE) Volatility Index (VIX), a volatility index may capture short-term mood sentiments and is an effective proxy instrument in behavioral finance. The VIX index displays the volatility expectations over the next 30 calendar days (Marquit and Curry 2023) and is considered a real-time monitoring index. As Siriopoulos and Fassas (2009) explain, the CBOE Volatility Index has been broadly accepted due to its advanced interpretative power in antithesis to historical volatility. Investors, fund managers, and traders use the VIX as a barometer to determine the market’s fear or level of risk when constructing their portfolios (Economou et al. 2018). The current study uses the VIX index based on the CBOE VIX methodology (Whaley 2000) to capture any ESG leaders’ investors’ bias in relation to volatility. Furthermore, the present study examines whether European and Global ESG leaders’ portfolios tilt towards the dispersion of returns by employing the cross-sectional absolute deviation of returns (CSAD), introduced by Chang et al. (2000), which is the most accepted method of return dispersion used in behavioral finance literature. In this way, we fill the gap in the literature and investigate any bias in the dispersion of returns on ESG leaders’ investments.

An interesting avenue of research arising from the above discussion will be to examine if there is any herding behavior in ESG leaders’ portfolios, motivating us to extend the behavioral finance literature. Herding behavior refers to imitating and following others’ investment decisions instead of following one’s beliefs and information. In the investment process, emotional biases can arise due to herding affecting investment decisions in selecting, buying, holding, or selling assets. Understanding and analyzing herding behavior is vital to the portfolio strategies of investors and fund managers. The cross-sectional absolute deviation of returns (CSAD) is the most common measure used to examine herding behavior, introduced by Chang et al. (2000). There is a small volume of existing studies investigating herding behavior using ESG data. Rubbaniy et al. (2021), Blondel (2022) and Gavrilakis and Floros (2023a) found evidence of herding behavior using ESG data, while Ciciretti et al. (2021) found anti-herding behavior.

The present study reports negative alphas on European ESG leaders’ portfolios by using the CAPM, FF3, and FF5 regression models. The negative abnormal returns of ESG leaders’ portfolios are homogeneous across the three ESG disclosures (environmental, social, and governance). Moreover, the evidence suggests that European ESG investing tilts towards large caps, and values stocks with robust operating profitability against aggressive strategies. We examine Global ESG leaders’ index-based portfolios in a robustness check, reporting the same outcomes but with insignificant importance in some loading factors like profitability and investment strategy. Furthermore, we document for the first time

that ESG leader's portfolios are biased toward volatility and herding behavior. These results shed light on risk–return tradeoffs and behavioral biases on the part of investors and finance professionals responding proactively and reducing uncertainty in the portfolio optimization process.

The current study contributes to the literature by providing valuable updates for investors and fund managers exposed to ESG leaders' assets. We show investors and professionals cannot achieve high-risk-adjusted returns when holding ESG leaders' stocks in the short run. This outcome is helpful during portfolio construction and diversification processes. The evidence that European ESG investing tilts towards large caps, and values stocks with robust operating profitability against aggressive strategies encourages their use in corporate financial decisions. In addition, understanding and exploring how the volatility and dispersion of returns interact with ESG investing is essential for constructing optimal portfolios. Finally, our work contributes by reporting for the first time that ESG leaders' portfolios are biased toward volatility and herding behavior, which is helpful for investors, fund managers, and analysts seeking to reduce risk-taking and hedge their portfolios.

The rest of this study is organized as follows: In Section 2, we review the literature and analyze alphas, volatility, and herding. In Section 3, the methodology is presented, while our results are reported in Section 4. Finally, Section 5 concludes the study.

2. Literature Review

2.1. ESG Abnormal Returns

Literature on ESG investment performance focuses on three main areas. The ESG investments result in positive returns, as Renneboog et al. (2008), Consolandi et al. (2009), Giese et al. (2019) and Yu (2022) claimed. Secondly, the ESG investments underperform, mainly due to lower diversification and short-term growth immolation, as supported by Hamilton et al. (1993), Carhart (1997), Bauer et al. (2005), Adler and Kritzman (2008) and Berlinger and Lovas (2015). Lastly, the neutrality argument is supported by Managi et al. (2012), Mollet and Ziegler (2014), Hartzmark and Sussman (2019), and Naffa and Fain (2022), who report insignificant abnormal returns, supporting the efficient market hypothesis (EMH) that all existing market information is reflected in asset prices, and alpha generation is impossible.

Following the work of Merton (1987), Heinkel et al. (2001), Zerbib (2020), and Pedersen et al. (2021), we presume that investors sensitive to ESG are averse to the assets of laggard ESG firms, and place increased demand, especially on the part of younger generations (Tao et al. 2020), on ESG leaders. Financial market participants have recently reported obligations (article 8, 9) to promote assets with high rather than low ESG scores (Chava 2014), and intentionally incorporate ESG leaders into their sustainable investment portfolios (Tao et al. 2020). In measuring Jensen's alpha abnormal returns, Zehir and Aybars (2020), using the CAPM regression model, reported that two ESG-based portfolio scores underestimate the market index, while, when using the FF3 model, the abnormal returns of the ESG bottom governance portfolio perform better than the rest of the market. In the same way, Teti et al. (2023), by employing the CAPM, FF3, and FF5 methodologies, found robust evidence that a bottom decile portfolio produces negative alphas. In a recent paper, Luo (2022) examined UK stocks from 2003 to 2020, and found that companies in the low ESG quintile perform better (value-weighted returns) by 0.513% per month than those in the high ESG quintile. In addition, Luo (2022) analyzed the returns of ESG portfolios constructed from stocks of the STOXX Europe 600 index, and found that the environmental, social, and governance premiums are significant. In their study, Naffa and Fain (2022) measured the performance of Global ESG assets by recommending a new FF5 factor approach, and deduced that ESG portfolios did not yield significant alphas during 2015–2019, verifying the neutrality argument of the literature. Similarly, Dhasmana et al. (2023) investigated the relationship between the MSCI ESG index and investor sentiment, and revealed the underperformance of the MSCI ESG index in India. Following these leads, we provide a route to the first hypothesis:

H1. *ESG leaders' portfolios produce negative alphas.*

2.2. *ESG, Volatility, and Dispersion of Returns*

Financial assets present volatility clustering and asymmetric behavior through financial uncertainty. Various volatility methods have been adopted to analyze the volatility behavior of asset returns (McKenzie and Mitchell 2002). The literature review here indicates some studies that have concentrated on the use of the volatility index (VIX) as a proxy for investor sentiment, with mixed results. The VIX index is known as the “investor fear gauge”, and captures investors’ expectations of market volatility (Durand et al. 2011).

In their paper, Durand et al. (2011) analyzed the role of the VIX and momentum in asset pricing, and concluded that the explanatory power of all factors in Fama and French’s three-factor model is enhanced by the inclusion of both the momentum and VIX factors. Chiang (2012), by using the GARCH methodology, examined the relationship among the S&P500, NASDAQ100, VIX, and VXN indices, and concluded that the VIX—fear index significantly affects the S&P500 Index. In the same way, López-Cabarcos et al. (2019) examined the effects of the VIX index on the indices of the S&P500 and the S&P500 environmental and social responsibility from 2015 to 2016, and found a negative correlation. Morales et al. (2019) reported a negative effect of the VIX index on Socially Responsible Investing (SRI) indices (DJSI World excluding alcohol, gambling, tobacco and arms index, Dow Jones Sustainability Index, MSCI KLD 400 Social index, and World USA subset index). Chen and Gao (2020) examined how three defined volatility risk factors derived from VIX may affect the pricing of assets by employing the three-factor Fama–French asset pricing model, and identified a significant relation of the risk factors to returns of individual portfolios. In their study, Öcal and Kamil (2021), using as risk indicators the VIX, the CDS, and FX volatility indices, revealed that stocks with higher ESG exposure in Germany, France, Indonesia, and Turkey are less affected by market crises than companies included in broad-based indices. Furthermore, Shaikh (2022) examined the relationship between sustainable investment and uncertainties, and found a negative relationship between the VIX index and Dow Jones Sustainability Indices from 2000 to 2017. Vergili and Celik (2023), in their study, reported a long-term cointegration relationship (with a negative coefficient) between the VIX fear index and the Dow Jones Sustainability Emerging Markets Index (DJSEMUP) by using monthly data from 2013 to 2020. Finally, by employing the Morgan Stanley Capital International (MSCI) ESG Leader index, Sabbaghi (2023) examined the volatility risk and provided seminal evidence of a slow response to news in emerging markets on the part of the selected index. A relatively new index, the EURO STOXX 50 Volatility Index (VSTOXX), provides the implied volatility over the next 30 days on the EURO STOXX 50 Index. Stanescu and Tunaru (2013), by adding VIX and VSTOXX volatility index futures to a portfolio, improved the return–risk profiles of portfolios, particularly during turbulent times. Cocozza et al. (2021) found the same behaviors in the VSTOXX and VIX indexes in their study.

According to Gleason et al. (2004) and Henker et al. (2004), dispersion measures the extent to which investors follow the market’s expectations. Ramadan (2015) stated that “if investors follow market expectations, then their returns will not deviate from the market return, and as a result, the dispersion level or variance between individuals’ return and market return will be zero”. Return dispersion (RD) continues to play a vital role in explaining the cross-sectional variation in expected returns, even if idiosyncratic volatility, market volatility, momentum factor, size, and book-to-market factors are included in different asset pricing models like CAPM, MVM, IVM, and FF-3 (Jiang 2010). Demirer and Jategaonkar (2013), in their study, observed a systematic conditional relation between dispersion and return even after controlling for market, size, and book-to-market factors by adding to the three Fama–French factors the return dispersion (RD) risk factor. Verousis and Voukelatos (2018) suggested that the cross-sectional dispersion (CSD) of stock returns is negatively associated with the investment, meaning that assets with high dispersion offer lower returns. Caoa et al. (2019) analyzed the relationship between the cross-sectional dispersion (CSD)

of returns and active fund performance in Australia, and concluded that outperformance occurs only for funds with high return dispersion. Last year, Qontigo presented a new index (EURO STOXX 50 realized dispersion index) to replicate a dispersion indicator on selected Eurozone blue chips “<https://qontigo.com/index/sx5edisp/>” (accessed on 7 July 2023)”. By using a modified version of the VIX methodology, the CBOE S&P500 Dispersion Index (DSPX) was used to measure the expected dispersion in the S&P500 over the next 30 calendar days “<https://www.cboe.com/us/indices/dispersion/>” (accessed on 7 July 2023)”. Chiang and Zheng (2010) suggested an approach to examining the dispersion measure: the cross-sectional absolute deviation of returns (CSAD). The main idea of this approach is that if the investors adhere to the same market expectations, the relationship between the cross-sectional deviation and market portfolio will be linear, and this means that the CSAD will decrease or increase, at least less than the relative rate of market return. The CSAD methodology is the primary indicator used for examining herding behavior in the next part of our study. This gives us ground to determine whether ESG leaders’ portfolios tilt towards the dispersion of returns by using the CSAD approach. Based on a related literature review, the current study explores the second hypothesis:

H2. *ESG leader’s portfolios tilt towards the volatility and dispersion of returns.*

2.3. ESG and Herding

An exciting and little-examined area of behavioral finance relates to herding behavior, and exploring whether herding bias can be more intense when volatility prevails in the ESG leader’s market. An increased demand for ESG investing (Benz et al. 2020) may guide investors and portfolio managers towards herding bias (Przychodzen et al. 2016; Rubbaniy et al. 2021). Investors and fund managers that herd imitate other investing actions without consideration of their own fundamental analysis or private information. The herding phenomenon has been empirically investigated in various markets, such as stocks, bonds, ETFs, REITs, ESG, cryptocurrency, and, lately, artificial intelligence.

A herding methodology based on the cross-dispersion of assets performed by Hwang and Salmon (2004) captured significant herding movements in the South Korean and US markets. Using data from the Pacific Basin stock market region, Chiang et al. (2013) pointed out a lower herding bias in positive volatility index (VIX) returns. In the same way, Economou et al. (2015) indicated herding behaviors on days with negative VIX index returns in the Euronext markets. Furthermore, by employing time-varying models, Arjoon and Bhatnagar (2017) captured herding effects across two frontier markets, the Trinidad and Tobago Stock Exchange, from 2001 to 2014. As Benz et al. (2020) claimed, investors, hedge funds, and fund managers display sustainable investment herding due to concerns about adhering to the market consensus, applying data (2007–2020) from the MSCI U.S.A. ESG leader index. Rubbaniy et al. (2021) detected herding behavior during bull and bear market periods. Furthermore, Fu and Wu (2021) used the Markov regime-switching model to capture herding bias in the Chinese stock market. Regarding sustainable investing, Blondel (2022) deduced that low-risk investors herd less than medium-risk investors by surveying 175 investors for two months in 2021, whereas in relation to asset picking, passive investors herd more than active investors. Gavrilakis and Floros (2023a) recently reported herding behavior in ESG investing in Portugal, Italy, and Greece during the COVID-19 outbreak. Finally, Ameye et al. (2023) found evidence of significant herding effects in 307 firms using artificial intelligence technology, and secondly, they concluded that uncertainty moderates herding.

Contrariness, an anti-herding behavior, was noticed in data concerning metal commodities futures by Babalos and Stavroyiannis (2015). The same result was reported by Rompotis (2018), who investigated a sample of 34 small-cap and 66 large-cap ETFs from 2012 to 2016. Moreover, Coskun et al. (2020) captured anti-herding behavior in 14 leading cryptocurrencies using GARCH, Time-Varying Markov-Switching (TV-MS), and CSAD methodologies. Using the same approach, Yarovaya et al. (2021) noticed no indications of

herding bias in the cryptocurrency market. Moreover, using a database of 10,456 Global ESG funds, Ciciretti et al. (2021) found anti-herding behavior in their sample from 2012 to 2018. During the COVID-19 pandemic, anti-herding predominated (Yang and Chuang 2023) due to market uncertainty. This literature has led us to develop and examine the third hypothesis:

H3. *ESG leaders' portfolios lean toward herding.*

3. Methodology and Data

3.1. Regression Analysis

In this paper, we expand the literature on ESG portfolio management by analyzing the abnormal returns of selected ESG leaders' portfolios and exploring factor loadings based on the well-known Capital Asset Pricing Model (CAPM), and the Fama–French three-factor (FF3) and five-factor (FF5) models. Furthermore, we extend this analysis by introducing two additional factors into the FF5 regression model: the volatility index (Vix_t) and the cross-sectional absolute deviation of returns ($CSAD_t$) indicator. The above methodologies enable us to develop greater awareness concerning excess ESG returns or the alphas, risk, and return dispersion.

The first model used in the regression analysis is the Capital Asset Pricing Model (CAPM), developed by Treynor (1961):

$$ESG_{it} - R_{ft} = a_{it} + \beta_1(MKT_t) + \varepsilon_{it} \quad (1)$$

where $ESG_{it} - R_{ft}$ is the excess return of European ESG leaders portfolios; R_{ft} is the one-year Treasury Bill rate; a_{it} is the Jensen's alpha; MKT_t is the market risk premium ($R_{mt} - R_{ft}$) on day t ; β_1 is the beta or the sensitivity of the ESG_{it} to the market, and ε_{it} is the error term. The CAPM is a suitable methodology for use in investigating the ESG performance with respect to the passive investment strategy.

The next model employed is the Fama and French (1993) three-factor model (FF3):

$$ESG_{it} - R_{ft} = a_{it} + \beta_1(MKT_t) + \beta_2(SMB_t) + \beta_3(HLM_t) + \varepsilon_{it} \quad (2)$$

where (SMB_t) and (HLM_t) are, respectively, the firm's size and value characteristics. SMB_t (small minus big) represents the size premium, meaning large-cap assets are expected to earn lower returns than small-cap assets (Zehir and Aybars 2020). HLM_t (high minus low) stands for the value premium; stocks with low book-to-market ratios are expected to underperform compared to those with high book-to-market ratios. The regression coefficients β_1 , β_2 , and β_3 explain the ESG leaders' portfolios' sensitivities to the pre-specified indicators.

Equation (3) is the Fama and French (2015) five-factor (FF5) model:

$$ESG_{it} - R_{ft} = a_{it} + \beta_1(MKT_t) + \beta_2(SMB_t) + \beta_3(HLM_t) + \beta_4(RMW_t) + \beta_5(CMA_t) + \varepsilon_{it} \quad (3)$$

where (RMW_t) and (CMA_t) are, respectively, returns for profitability and investment factors. RMW_t (robust minus weak) relates to the profitability premium, meaning stocks with weak operating profitability are foreseen to underperform stocks with robust operating profitability. CMA_t (conservative minus aggressive) is the return difference between stocks that invest conservatively minus those that invest aggressively. The regression coefficients β_4 and β_5 are the ESG leaders' portfolios' sensitivities to profitability and investment factors.

The effects of volatility and herding are crucial disclosures when determining abnormal portfolio returns. We use the respective implied volatility index (Vix_t) and the cross-sectional absolute deviation of returns ($CSAD_t$) factor to estimate the impact of the volatility and dispersion of returns on ESG leaders' portfolio performances. Therefore, we extend the FF5 model to the FF7 model, augmented by addition of the Vix_t indicator and the herding $CSAD_t$ indicator, as follows:

$$ESG_{it} - R_{ft} = a_{it} + \beta_1(MKT_t) + \beta_2(SMB_t) + \beta_3(HLM_t) + \beta_4(RMW_t) + \beta_5(CMA_t) + \beta_6(Vix_t) + \beta_7(CSAD_t) + \varepsilon_{it} \quad (4)$$

We applied the traditional OLS with Newey and West's (1987) standard errors to evaluate the coefficients.

The value of MKT_t is the market risk premium $(R_{mt} - R_{ft})$, or the excess return on the market, and is calculated as the value-weighted return on all NYSE, AMEX, and NASDAQ stocks minus the one-month Treasury bill rate (R_{ft}). The values MKT_t , SMB_t , HLM_t , RMW_t , and CMA_t in the regression models were derived from the Kenneth R. French Library (2023).

3.2. Herding Analysis

In checking the robustness for herding, we employed the cross-sectional absolute deviation of returns (CSAD), introduced by Chang et al. (2000), which is the most accepted method of return dispersion in the behavioral finance literature:

$$CSAD_t = \frac{1}{N} \sum_{i=1}^N |R_{i,t} - R_{m,t}| \quad (5)$$

where $CSAD_t$ is the return dispersion of assets on day t , $R_{i,t}$ is the return of the i th asset on day t , $R_{m,t}$ represents the market return on day t , and N is the number of assets in the ESG index employed. To capture any herding behavior, we used the following model set out by Chang et al. (2000):

$$CSAD_t = \alpha + \gamma_1 |R_{m,t}| + \gamma_2 R_{m,t}^2 + \varepsilon_t \quad (6)$$

where $R_{m,t}^2$ is used for checking the non-linearity of the relationship, and γ_1 will be positive and γ_2 will be equal to zero in the absence of any herding. A significant and negative value of the parameter γ_2 suggests a herding bias since it verifies "that during a market disturbance, a nonlinear negative relationship exists between return dispersion and $R_{m,t}^2$ " (Rompotis 2018). Concerning herding asymmetries, we follow the approaches of Chang et al. (2000), Chiang and Zheng (2010), Economou et al. (2011), and Philippas et al. (2013). To evaluate the asymmetric behavior of market return dispersion, we apply Equation (7):

$$CSAD_t = a + \gamma_1 D^{up} |R_{m,t}| + \gamma_2 (1 - D^{up}) |R_{m,t}| + \gamma_3 D^{up} (R_{m,t})^2 + \gamma_4 (1 - D^{up}) (R_{m,t})^2 + \varepsilon_t \quad (7)$$

D^{up} is a dummy variable taking a value of 1 when there are positive market returns and 0 on days with negative market returns. We run Equation (7) to estimate any herding behavior during the development of a market. In the absence of herding, both parameters γ_1 and $\gamma_2 > 0$. If γ_3 and γ_4 are negative and $\gamma_3 > \gamma_4$, then herding is more evident during negative market returns. The above models allow us to capture any herding behavior or bias in the ESG leaders' portfolios.

4. Results and Discussion

This section presents the results of European ESG leaders' portfolios and the outcomes of robustness tests of Global ESG leaders' index-based portfolios. Tables 1 and 2 describe the indices of selected European and Global ESG leaders. Figure 1 illustrates that the cumulative return of the benchmark index (STOXX Europe 600) significantly exceeded those of the European ESG leaders' pillars, ending in a full-sample cumulative return of 76.01% from 1 June 2012 to 15 July 2022. The performance of the STOXX Europe ESG leaders' portfolio (9.92%) is the most notable. The environmental pillar reported a 22.22% return, the social pillar displayed a 16.07% return, and the governance pillar resulted in a cumulative return of 23.20%. Figure 2 depicts the market returns of the STOXX Global 1800 index and the Global ESG pillar leaders' indices. The cumulative return of the global

benchmark index was 170.3% from 1 June 2012 to 15 July 2022. The cumulative return of STOXX Global ESG leaders was 105.42%, close to the social pillar, with 101.5% return, and the governance pillar, with 98.35% return. The environmental pillar showed a 119.41% cumulative return. The above findings show considerably higher returns on Global ESG leaders’ investments than European ESG leaders’ investments.

Table 1. Descriptions of the selected European ESG leaders’ indices (Qontigo 2023).

ESG and the Benchmark Index	Index Purpose/Descriptions from Factsheet (Qontigo 2023)
Stoxx Europe ESG Environmental Leaders Index	The STOXX Europe ESG Environmental Leaders Index aims to measure the performance of portfolios of equity that have been selected with reference to their business interests and policies toward environmental issues, and which are designed to provide exposure to certain variables known as factors; these factors can assist in the explanation of market performance. A detailed ESG report describing how these factors are applied to this index is available on the STOXX website, “ www.stoxx.com/resources (accessed on 7 July 2023)”.
Stoxx Europe ESG Social Leaders Index	The STOXX Europe ESG Social Leaders Index aims to measure the performance of portfolios of securities that have been selected with reference to their business interests and policies toward social issues, and which are designed to provide exposure to certain variables known as factors; these factors can assist in the explanation of market performance. A detailed ESG report describing how these factors are applied to this index is available on the STOXX website, “ www.stoxx.com/resources (accessed on 7 July 2023)”.
Stoxx Europe ESG Governance Leaders Index	The STOXX Europe ESG Governance Leaders Index aims to measure the performance of portfolios of equity that have been selected with reference to their business interests and policies toward governance issues, and which are designed to provide exposure to certain variables known as factors; these factors can assist in the explanation of market performance. A detailed ESG report describing how these factors are applied to this index is available on the STOXX website, “ www.stoxx.com/resources (accessed on 7 July 2023)”.
Stoxx Europe ESG Leaders Select 30 Index	The STOXX ESG Leaders Select 30 Index aims to measure the performance of portfolios of securities that have been selected with reference to their business interests and policies toward environmental, social, and governance issues. This index includes benchmarks in which the portfolios are weighted by free-float market capitalization; portfolios may also be weighted to reduce risk using volatility or variance weighting schemes. “ www.stoxx.com/resources (accessed on 7 July 2023)”.
Stoxx Europe 600 Index (Benchmark index)	The STOXX Europe 600 index is a stock index of European stocks designed by STOXX Ltd. It has 600 components representing large, mid, and small capitalization companies from 17 European countries, covering approximately 90% of the free-float market cap. “ https://qontigo.com/index/sxxp/ (accessed on 7 July 2023)”

Source: Qontigo (2023). Available at: “<https://qontigo.com/> (accessed on 7 July 2023)”.

Table 2. Descriptions of the selected Global ESG Leaders indices (Qontigo 2023).

ESG and the Benchmark Index	Index Purpose/Descriptions from Factsheet (Qontigo 2023)
Stoxx Global ESG Environmental Leaders Index	The STOXX Global ESG Environmental Leaders index provides access to global environmental leaders through quantitative selection. The sustainability data in environmental areas are supplied by Sustainalytics ("https://www.sustainalytics.com/esg-data (accessed on 7 July 2023)"). The index follows a bottom-up approach based on the company's ESG scores. The system ranges from 0 to 100 points and is applied to the environmental pillar. Index components are weighted according to their ESG scores "www.stoxx.com/indices/rulebooks.html (accessed on 7 July 2023)".
Stoxx Global ESG Social Leaders Index	The STOXX Global ESG Social Leaders index provides access to global social leaders through quantitative selection. The sustainability data in social areas are supplied by Sustainalytics ("https://www.sustainalytics.com/esg-data (accessed on 7 July 2023)"). The index follows a bottom-up approach based on the company's ESG scores. The system ranges from 0 to 100 points and is applied to the social pillar. Index components are weighted according to their ESG scores "www.stoxx.com/indices/rulebooks.html (accessed on 7 July 2023)".
Stoxx Global ESG Governance Leaders Index	The STOXX Global ESG Governance Leaders index provides access to global governance leaders through quantitative selection. The sustainability data in governance areas are supplied by Sustainalytics (https://www.sustainalytics.com/esg-data). The index follows a bottom-up approach based on the company's ESG scores. The system ranges from 0 to 100 points and is applied to the governance pillar. Index components are weighted according to their ESG scores "www.stoxx.com/indices/rulebooks.html (accessed on 7 July 2023)".
Stoxx Global ESG Leaders Index	The STOXX Global ESG Leaders' indices consist of one broad and three specialized indexes for environmental, social, and governance pillars. The three specialized indices form the broad STOXX Global ESG Leaders Index. The indices provide access to global sustainability leaders through quantitative selection. The sustainability data in environmental, social, and governance areas are supplied by Sustainalytics ("https://www.sustainalytics.com/esg-data (accessed on 7 July 2023)").
Stoxx Global 1800 Index (Benchmark index)	The STOXX Global 1800 Index contains 600 European, 600 North American, and 600 Asia/Pacific region stocks represented by the STOXX Europe 600 Index, the STOXX North America 600 Index, and the STOXX Asia/Pacific 600 Index. "https://qontigo.com/index/sxw1e/ (accessed on 7 July 2023)".

Source: Qontigo (2023). Available at: ["https://qontigo.com/](https://qontigo.com/) (accessed on 7 July 2023)".

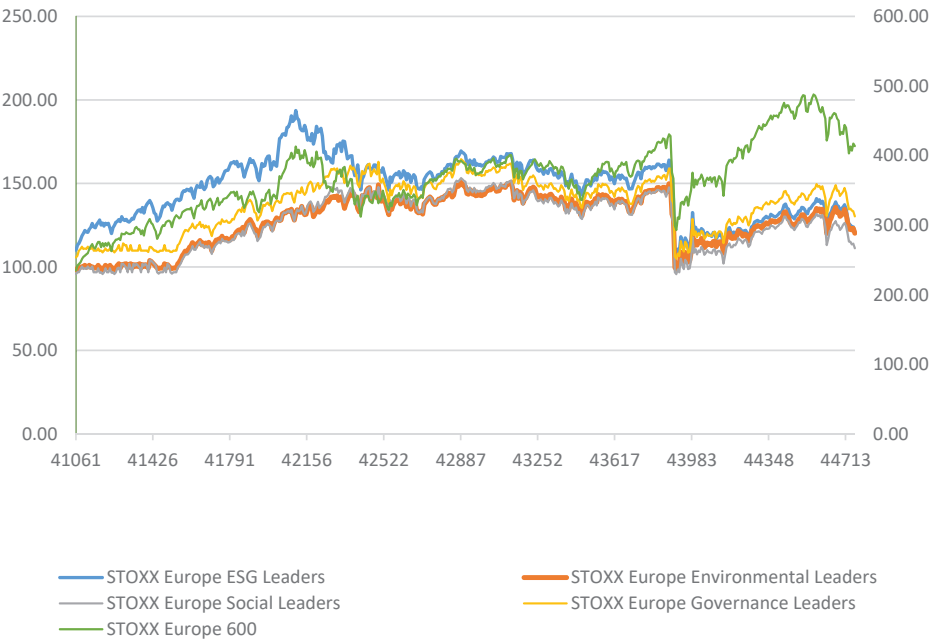


Figure 1. Cumulative returns of Europe ESG pillars leaders related to the benchmark (STOXX Europe 600).

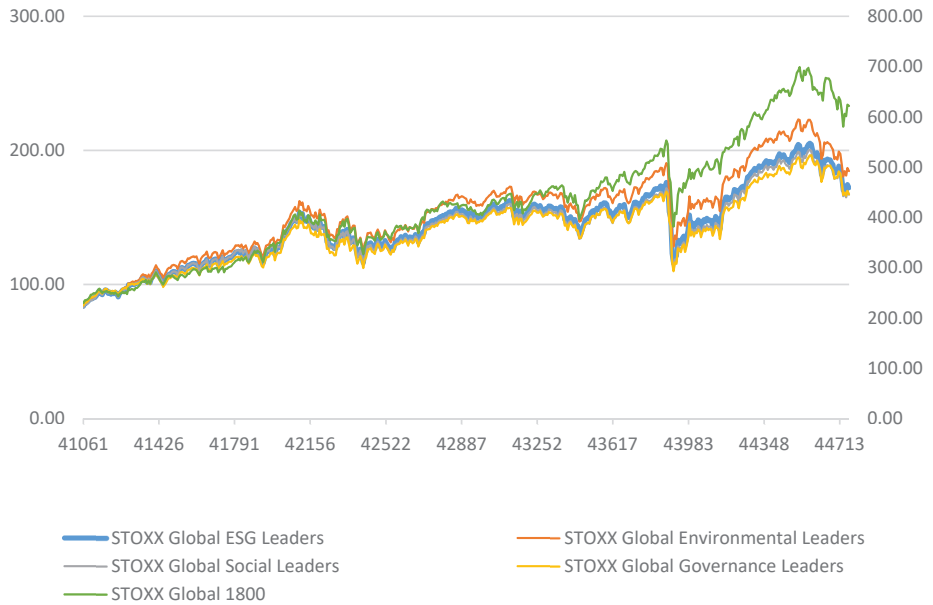


Figure 2. Cumulative returns of Global ESG pillars leaders related to the benchmark (STOXX Global 1800) index.

Tables 3 and 4 show the correlation matrices of selected European and Global ESG pillars, along with the volatility and dispersion of returns factor. All indices are highly positively correlated, while the series is less positively correlated with the volatility index and slightly negatively correlated with the dispersion of returns index, implying that

when the dispersion of returns grows, the ESG leaders’ performances decline, which strongly adheres to the risk–return theory. Tables 5 and 6 depict the descriptive statistics of European and Global ESG leaders’ portfolios. The results illustrate, on average, a significant dispersion across all ESG leaders’ indices. The annual mean standard deviation (volatility) suggests that ESG leaders’ investing indicates low risk and negative asymmetry. The summary statistics show the daily returns’ negative asymmetry (negative skewness values). The distributions of ESG indices are leptokurtic, as the kurtosis values are above 12, indicating the presence of fat tails in the series. Kurtosis is only valid when used in connection with standard deviation. ESG indices have high kurtosis, which is unsuitable for investment strategies as the returns are close to the mean, but the overall volatility of the ESG indices is low, which should be considered. The distributions are non-normal for all series as the Jarque–Bera statistics show significant *p*-values, indicating the departure from normality and the arrival of volatility clustering. In some cases, we apply logarithms into our series to overcome non-normality. The Augmented Dickey–Fuller (ADF) test results have significant *p*-values. As a result, we reject the H0 hypothesis (the series have a unit root problem) and accept the stationarity of the series under study. This finding aligns with that of Gavrilakis and Floros (2023b), who concluded that the ESG/thematic market does not have a random walk (non-normal) distribution and that predicting a long-term relationship between the series’ elements is feasible. In the same study, the authors employed several Copula models to reproduce asymmetries in the asymptotic tail dependence of stationary ESG/thematic investment. Copulas have been used extensively in quantitative finance to capture tail risk and, lately, in applications for portfolio optimization (Patton 2012; Dewick and Liu 2022; Nagler et al. 2022). As suggested by many studies on ESG investing that support the null hypothesis of stationarity (Jain et al. 2020; Górka and Kuziak 2021; Ouchen 2022; Erol et al. 2023), we now proceed by applying an econometric analysis to the reference series.

Table 3. Correlation matrix of selected European ESG Leaders indices.

	Stoxx Europe ESG Environmental Leaders Index	Stoxx Europe ESG Social Leaders Index	Stoxx Europe ESG Governance Leaders Index	Stoxx Europe ESG Leaders Select 30 Index	Stoxx Europe 600 Index	VIX _t	CSAD _t
Stoxx Europe ESG Environmental Leaders Index	1						
Stoxx Europe ESG Social Leaders Index	0.9826	1					
Stoxx Europe ESG Governance Leaders Index	0.9868	0.9839	1				
Stoxx Europe ESG Leaders Select 30 Index	0.7528	0.7577	0.7532	1			
Stoxx Europe 600 Index	0.7474	0.7433	0.7410	0.7507	1		
VIX _t	0.1389	0.1377	0.1392	0.1504	0.1846	1	
CSAD _t	−0.0659	−0.0665	−0.0635	−0.0735	−0.0330	−0.0203	1

Table 4. Correlation matrix of selected Global ESG Leaders indices.

	Stoxx Global ESG Environmental Leaders Index	Stoxx Global ESG Social Leaders Index	Stoxx Global ESG Governance Leaders Index	Stoxx Global ESG Leaders Index	Stoxx Global 1800 Index	VIXt	CSADt
Stoxx Global ESG Environmental Leaders Index	1						
Stoxx Global ESG Social Leaders Index	0.9938	1					
Stoxx Global ESG Governance Leaders Index	0.9908	0.9950	1				
Stoxx Global ESG Leaders Select Index	0.9967	0.9981	0.9972	1			
Stoxx Global 1800 Index	0.6572	0.6453	0.6310	0.6463	1		
VIXt	0.1906	0.1860	0.1892	0.1900	0.0746	1	
CSADt	−0.1338	−0.1331	−0.1357	−0.1347	−0.0913	−0.0611	1

Table 5. Descriptive statistics of European ESG Leaders indices.

Statistic	Stoxx Europe ESG Environmental Leaders Index	Stoxx Europe ESG Social Leaders Index	Stoxx Europe ESG Governance Leaders Index	Stoxx Europe ESG Leaders Select 30 Index	Stoxx Europe 600 Index
Min	−0.1381	−0.1419	−0.1383	−0.1460	0.1147
Max	0.0778	0.0766	0.0793	0.0860	0.0840
Mean	6.8668×10^5	4.4887×10^5	6.1835×10^5	$−2.7927 \times 10^5$	0.0002
SD	0.0089	0.0091	0.0091	0.0106	0.0103
Skewness	−2.2557	−2.1594	−2.2102	−1.3565	−0.8440
Kurtosis	34.012	33.498	32.635	20.774	11.112
JB test	124,832 ***	120,966 ***	115,009 ***	46,564 ***	13,400.6 ***
ADF test	−13.4561 ***	−12.4959 ***	−13.236 ***	−18.4363 ***	−18.1139 ***

Normality test based on skewness, kurtosis values, Jarque–Bera test, and Augmented Dickey–Fuller test, *** $p < 0.01$.

Table 6. Descriptive statistics of Global ESG Leaders indices.

Statistic	Stoxx Global ESG Environmental Leaders Index	Stoxx Global ESG Social Leaders Index	Stoxx Global ESG Governance Leaders Index	Stoxx Global ESG Leaders Index	Stoxx Global 1800 Index
Min	−0.1069	−0.1125	−0.1131	−0.1108	−0.0983
Max	0.0881	0.0959	0.0916	0.0919	0.0826
Mean	0.0002	0.0001	0.0001	0.0002	0.0004
SD	0.0098	0.0102	0.0101	0.0100	0.0094
Skewness	−0.7180	−0.7054	−0.7848	−0.7432	−0.8955
Kurtosis	12.384	13.265	13.270	13.066	14.453
JB test	16,488.4 ***	18,878.4 ***	18,935.3 ***	18,345.6 ***	22,489.9 ***
ADF test	−16.1624 ***	−16.0008 ***	−16.0273 ***	−16.0829 ***	−13.1973 ***

Normality test based on skewness, kurtosis values, Jarque–Bera test, and Augmented Dickey–Fuller test, *** $p < 0.01$.

Table 7 illustrates the OLS regression results (with HAC standard errors) of European ESG Leaders’ index-based portfolios. The validity of the OLS model was tested based on multicollinearity tests detecting values below ten. Therefore, there is no evidence of a multicollinearity problem. The results of all factor models suggest that all the ESG leaders’ portfolios underperformed (with negative risk-adjusted abnormal returns) against the market. We confirm the findings of Mollet and Ziegler’s (2014) study, which identified insignificant abnormal returns on three different portfolios (“MSCI sustainability leaders”, “Sustainability leaders”, and “Other MSCI firms”) for Europe and the US from 1998 to 2009. Irrespective of our sample, the ESG portfolios present medium betas (β), indicating a medium level of systematic risk. The ESG portfolios depict a negative loading on SMB_t , implying a tilt toward large-cap firms. The factor loadings for the determinant HLM_t are broadly significantly positive in all series, which indicates a bias towards value stocks. The ESG portfolios report a favorable loading on RMW_t , indicating a tilt toward robust operating profitability. The statistically significant and positive CMA_t exposure suggests that the ESG portfolios include firms with conservative investment strategies, usually associated with low future returns. The factor loadings are homogeneous across the environmental, social, and governance components. The ESG premium remains statistically significant after adjusting for the CAPM, FF3, FF5, and augmented FF7 models. This finding supports the results in the literature that FF3 and FF5 effectively explain ESG leaders’ performance concerning market returns (Zaremba and Czapkiewicz 2017; Guo et al. 2017). Furthermore, the results confirm that our augmented FF7 model significantly explains all the factor loadings. Table 7 also reports that the fear or volatility index (Vix_t) is significant and positive in all ESG indices, indicating a volatility bias. We verify the findings of the study of Górka and Kuziak (2021), who confirmed a higher dependence on the volatility of selected ESG indices from 2007 until 2019. Finally, the $CSAD_t$ factor is significant and negative in European ESG Leaders’ portfolios, suggesting a dispersion of return bias. To check the robustness of our results, we rerun the FF7 OLS model by adding the VSTOXX index instead of VIX, ending up with the same results (see Table 8) but with minor differences in the importance of loading factors ($CSAD_t$).

Table 7. Regression equation table for European ESG Leaders index-based portfolios. The regression results of STOXX Europe index-based ESG pillars against CAPM, FF3, FF5, and FF7 (including VIX_t and $CSAD_t$ in the FF5 model). Daily price data were obtained from the Refinitive Eikon database (2022) from 1 June 2012 to 14 July 2022. The model FF7 takes the form: $ESG_{it} - R_{ft} = a_{it} + \beta_1(MKT_t) + \beta_2(SMB_t) + \beta_3(HLM_t) + \beta_4(RMW_t) + \beta_5(CMA_t) + \beta_5(Vix_t) + \beta_5(CSAD_t) + \varepsilon_{it}$.

Model		Stoxx Europe ESG Environmental Leaders	Stoxx Europe ESG Social Leaders	Stoxx Europe ESG Governance Leaders	Stoxx Europe ESG Leaders Select 30	Stoxx Europe 600
CAPM	α	−0.0021 ***	−0.0022 ***	−0.0021 ***	−0.0023 ***	−0.0020 ***
	β	0.00601 ***	0.0062 ***	0.0060 ***	0.0092 ***	0.0083 ***
	R^2	0.4376	0.4447	0.43141	0.748	0.6467
FF3	α	−0.0021 ***	−0.0021 ***	−0.0021 ***	−0.0021 ***	−0.0020 ***
	β	0.0053 ***	0.0056 ***	0.0054 ***	0.0059 ***	0.0074 ***
	SMB_t	−0.0021 ***	−0.0017 ***	−0.0020 ***	−0.0046 ***	−0.0044 ***
	HML_t	0.00151 ***	0.0018 ***	0.0017 ***	0.0009 ***	−0.0004 *
	R^2	0.453	0.460	0.4486	0.5360	0.6687

Table 7. Cont.

Model		Stoxx Europe ESG Environmental Leaders	Stoxx Europe ESG Social Leaders	Stoxx Europe ESG Governance Leaders	Stoxx Europe ESG Leaders Select 30	Stoxx Europe 600
FF5	α	−0.0022 ***	−0.0022 ***	−0.0022 ***	−0.0022 ***	−0.0020 ***
	β	0.0056 ***	0.0058 ***	0.0056 ***	0.0062 ***	0.0075 ***
	SMB_t	−0.0018 ***	−0.0014 ***	−0.0016 ***	−0.0041 ***	−0.0043 ***
	HML_t	0.0023 ***	0.0028 ***	0.0027 ***	0.0016 ***	−0.0003
	RMW_t	0.0042 ***	0.0049 ***	0.0049 ***	0.0050 ***	0.0005
	CMA_t	0.0029 ***	0.0029 ***	0.0031 ***	0.0043 ***	0.0009
	R^2	0.4635	0.4731	0.4615	0.5501	0.6694
FF7	α	−0.0018	−0.0019 ***	−0.0019 ***	−0.0016 ***	−0.0023 ***
	β	0.0054 ***	0.0057 ***	0.0055 ***	0.0060 ***	0.0073 ***
	SMB_t	−0.0020 ***	−0.0016 ***	−0.0019 ***	−0.0046 ***	−0.0046 ***
	HML_t	0.0024 ***	0.0030 ***	0.0029 ***	0.0019 ***	$−4.16848 \times 10^6$
	RMW_t	0.0044 ***	0.0050 ***	0.0050 ***	0.0052 ***	0.0008
	CMA_t	0.0027 ***	0.0028 ***	0.0029 ***	0.0040 ***	0.0006
	VIX_t	0.0053 ***	0.0052 ***	0.0057 ***	0.0095 ***	0.0098 ***
	$CSAD_t$	−0.0376 *	−0.0371 *	−0.0333	−0.0686 **	0.0300
	R^2	0.4653	0.4746	0.4634	0.5555	0.6742

Note: ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. The risk-free rate (R_{ft}) and the values MKT_t , SMB_t , HLM_t , RMW_t , and CMA_t for the regression models were derived from the data in the Kenneth R. French Library. Alpha (a) and beta (b) are expressed in basis units (e.g., an alpha of −0.0021 is −0.21%, while a beta of 0.00601 is a 0.60 beta score). The R^2 is adjusted and describes the goodness of fit of the model.

Table 8. Regression equation table for European ESG leaders’ index-based portfolios. The regression results of STOXX Europe index-based ESG pillars against CAPM, FF3, FF5, and FF7 (including $VSTOXX_t$ and $CSAD_t$ in the FF5 model). Daily price data were obtained from the Refinitive Eikon database (2022) from 1 June 2012 to 14 July 2022. The model FF7 takes the form: $ESG_{it} - R_{ft} = a_{it} + \beta_1(MKT_t) + \beta_2(SMB_t) + \beta_3(HLM_t) + \beta_4(RMW_t) + \beta_5(CMA_t) + \beta_5(VSTOXX_t) + \beta_5(CSAD_t) + \epsilon_{it}$.

Model		Stoxx Europe ESG Environmental Leaders	Stoxx Europe ESG Social Leaders	Stoxx Europe ESG Governance Leaders	Stoxx Europe ESG Leaders Select 30	Stoxx Europe 600
FF7	α	−0.0017 ***	−0.0018 ***	−0.0018 ***	−0.0021 ***	−0.0022 ***
	β	0.0042 ***	0.0045 ***	0.0044 ***	0.0093 ***	0.0084 ***
	SMB_t	−0.0013 ***	−0.0010 ***	−0.0012 ***	−0.0004 *	−0.0035 ***
	HML_t	0.0027 ***	0.0032 ***	0.0031 ***	0.0016 ***	0.00038
	RMW_t	0.0048 ***	0.0054 ***	0.0054 ***	0.0033 ***	0.0013 **
	CMA_t	0.0025 ***	0.0026 ***	0.0027 ***	0.0037 ***	0.0003
	$VSTOXX_t$	0.0011 ***	0.0010 ***	−0.0010 ***	0.0001 ***	−0.0017 ***
	$CSAD_t$	−0.0423	−0.0416 *	−0.0376 *	0.0309 *	0.0426
	R^2	0.4837	0.4902	0.4780	0.7640	0.7086

Note: ***, **, * denote statistical significance at the 1%, 5%, and 10% levels, respectively. The risk-free rate (R_{ft}) and the values MKT_t , SMB_t , HLM_t , RMW_t , and CMA_t for the regression models were derived from the Kenneth R. French data Library. Alpha (a) and beta (b) are expressed in basis units (e.g., an alpha of −0.0017 is −0.17%, while a beta of 0.0042 is a 0.42 beta score). The R^2 is adjusted and describes the goodness of fit of the model.

We employed the cross-sectional absolute deviation of returns (Chang et al. 2000) methodology to enhance our results concerning the dispersion of returns bias and capture any herding effect. Table 9 summarizes our empirical results concerning herding existence in ESG leaders' portfolios. The benchmark methodology of herding presented by Chang et al. (2000) is applied in model I. Parameter γ_2 is significant and negative, implying the herding effect. This critical finding aligns with those of Benz et al. (2020), Blondel (2022) and Gavrilakis and Floros (2023a), who confirmed herding behavior in their studies related to ESG investing. ESG leaders' investors are involved in herd behavior during flat markets, which may result in market inefficiency and less diversified portfolios. Finally, our outcomes do not indicate herding behavior for up- or down-market days, as coefficients γ_1 and $\gamma_2 > 0$. This is contrary to the findings of Rubbaniy et al. (2021), who captured herding behavior during bull and bear market periods. This anti-herding behavior probably indicates that highly ESG-scoring assets conduce to market efficiency by lowering the probability of forming a financial bubble. Figure 3 presents the effects of herding on STOXX Europe ESG leaders' portfolios. There is a negative correlation between the dispersion of returns and the performance of ESG leaders' indices, which means that a 1-unit change in ESG leaders' returns leads to a -0.79 -unit change in the dispersion of returns (see Figure 3 and Table 9). We agree with Verousis and Voukelatos (2018), who found a negative relation between the dispersion of returns and investment performance.



Figure 3. Herding in STOXX Europe ESG leaders.

To check the robustness of our empirical results for European ESG leaders, we ran the same tests but with Global ESG leaders' portfolio data. Table 10 reports the OLS (with HAC standard errors) robustness regression results for Global ESG leaders' index-based portfolios. The validity of the OLS model was tested based on multicollinearity tests detecting values below ten. Therefore, there is no evidence of a multicollinearity problem. The findings of all the factor models used suggest that the Global ESG leaders' portfolios underperformed in the market from 1 June 2012 to 15 July 2022. This argument confirms the findings of Hartzmark and Sussman (2019), who found causal evidence that investors value sustainability, but no evidence of positive effects on performance. In addition, we partly confirm the findings of Auer and Schuhmacher (2016), who argued that high- or low-rated ESG stock selection does not provide investors with superior returns, while the returns for Asia-Pacific and the US are similar to ESG performance, but for European markets, they are lower. Moreover, all the ESG portfolios show medium betas (β), indicating a

medium level of systematic risk. The Global ESG portfolios presented a positive loading on SMB_t , but this was not significant, except for in the social portfolio, implying a tilt towards small firms. The factor loadings for the determinant HLM_t are significantly positive in all series, which indicates a bias towards value stocks. The RMW_t and CMA_t exposure are not statistically significant in all factor models. In summary, the factor loadings are similar across the environmental, social, and governance components. According to our results, the volatility index (Vix_t) is significant and positive in all portfolios, indicating a volatility bias. This finding agrees with those of Nishant et al. (2022), who argued that global companies with better ESG scores are more resilient in high-volatility environments. Furthermore, the $CSAD_t$ indicator is significant and negative in all portfolios, suggesting a dispersion of returns bias.

Once more, we employed the cross-sectional absolute deviation of returns to attest to the dispersion of returns bias and capture any herding effects (Chang et al. 2000). Model I in Table 11 shows evidence of herding, since coefficient γ_2 is statistically significant and negative, while our empirical results do not indicate herding behavior toward the ESG leaders' returns on up- or down-market days, as coefficients γ_1 and $\gamma_2 > 0$. Overall, our findings align with those of Rubbaniy et al. (2021) and Fu and Wu (2021), who confirmed herding behavior in their studies related to Global ESG investing. We do not confirm the findings of Ciciretti et al. (2021), who captured no herding behavior in 10,456 global ESG funds. Figure 4 presents the effects of herding on the STOXX Global ESG leaders' portfolio. There is a negative correlation between the dispersion of returns and the performance of ESG leader's indices, which means that a 1-unit change in ESG leaders' returns leads to a -0.13 -unit change in the dispersion of returns (see Figure 4 and Table 11). We contradict the conclusions of Caoa et al. (2019), who found a positive relation between fund performance and return dispersion.

Table 9. Results of herding models for Europe. This table indicates the results concerning the estimated coefficients of 2 fixed-effect herding models for STOXX Europe ESG Leaders Select 30 using a dataset from 1 June 2012 to 14 July 2022. Newey and West's (1987) standard errors were used. Model I: $CSAD_t = a + \gamma_1|R_{m,t}| + \gamma_2(R_{m,t})^2 + \varepsilon_t$; Model II: $CSAD_t = a + \gamma_1 D^{up}|R_{m,t}| + \gamma_2(1 - D^{up})|R_{m,t}| + \gamma_3 D^{up}(R_{m,t})^2 + \gamma_4(1 - D^{up})(R_{m,t})^2 + \varepsilon_t$.

Parameter	Model I	Model II
R-squared	0.3867	0.3954
Adj. R-squared	0.3862	0.3945
Log-likelihood	11,077.03	11,095.34
a	0.0068 ***	0.0070 ***
γ_1	0.3540 ***	0.2893 ***
γ_2	−0.7990 **	0.3468 ***
γ_3		1.8071 ***
γ_4		−0.9875 ***
Herding ($\gamma_2 < 0$)		Anti-herding ($\gamma_3 > 0, \gamma_4 < 0$)

Note: ***, **, denote statistical significance at the 1% and 5% levels, respectively.

Table 10. Robustness regression equation table for Global ESG leaders index-based portfolios. The regression results of STOXX Global index-based ESG pillars against CAPM, FF3, FF5, and FF7 (including VIX_t and $CSAD_t$ in the FF5 model). Daily price data were obtained from Refinitive Eikon database (2022) from 1 June 2012 to 14 July 2022. The model FF7 takes the form: $ESG_{it} - R_{ft} = a_{it} + \beta_1(MKT_t) + \beta_2(SMB_t) + \beta_3(HLM_t) + \beta_4(RMW_t) + \beta_5(CMA_t) + \beta_5(Vix_t) + \beta_5(CSAD_t) + \epsilon_{it}$.

Model		Stoxx Global ESG Environmental Leaders	Stoxx Global ESG Social Leaders	Stoxx Global ESG Governance Leaders	Stoxx Global ESG Leaders	Stoxx Global 1800 Index
CAPM	α	−0.0022 ***	−0.0022 ***	−0.0022 ***	−0.0022 ***	−0.0021 ***
	β	0.0059 ***	0.0061 ***	0.0059 ***	0.0060 ***	0.0075 **
	R^2	0.3963	0.3897	0.3663	0.3854	0.6890
FF3	α	−0.0022 ***	−0.0022 ***	−0.0022 ***	−0.0022 ***	−0.0021 ***
	β	0.0059 ***	0.0061 ***	0.0059 ***	0.0059 ***	0.0076 ***
	SMB_t	0.0003	0.0005 *	0.0003	0.0004	−0.0006 ***
	HML_t	0.0025 ***	0.0030 ***	0.0031 ***	0.0029 ***	0.0006 ***
	R^2	0.4359	0.4444	0.4247	0.4365	0.6934
FF5	α	−0.0022 ***	−0.0022 ***	−0.0022 ***	−0.002 ***	−0.0021 ***
	β	0.0059 ***	0.0060 ***	0.0058 ***	0.0059 ***	0.0076 ***
	SMB_t	0.0004	0.0005 *	0.0003	0.0004	−0.0006 ***
	HML_t	0.0025 ***	0.0030 ***	0.0031 ***	0.0029 ***	0.0008
	RMW_t	−0.0003	−0.0005	−0.0005	−0.0004	2.41469×10^5
	CMA_t	−0.0002	1.56867×10^5	0.0001	-5.11484×10^6	−0.0002
	R^2	0.4360	0.4448	0.4250	0.4367	0.6927
FF7	α	−0.0016 ***	−0.0016 ***	−0.0016 ***	−0.0016 ***	−0.0021 ***
	β	0.0059 ***	0.0061 ***	0.0059 ***	0.0059 ***	0.0076 ***
	SMB_t	0.0003	0.0005 *	0.0003	0.0004	−0.0006 ***
	HML_t	0.0025 ***	0.0029 ***	0.0030 ***	0.0028 ***	0.0007 ***
	RMW_t	−0.0002	−0.0004	−0.0004	−0.0003	6.75079×10^5
	CMA_t	−0.0002	-3.72079×10^5	0.0001	-5.96169×10^5	−0.0002
	VIX_t	0.0249 ***	0.0250 ***	0.0252 ***	0.0251 ***	0.0117 ***
	$CSAD_t$	−0.1040 ***	−0.1075 ***	−0.1139 ***	−0.1085 ***	−0.0063
	R^2	0.4766	0.4836	0.4649	0.4770	0.7033

Note: ***, **, * denote statistical significance at the 1%, 5%, and 10% levels, respectively. The risk-free rate (R_{ft}) and the values MKT_t , SMB_t , HLM_t , RMW_t , and CMA_t for the regression models were derived from the Kenneth R. French data Library. Alpha (a) and beta (b) are expressed in basis units (e.g., an alpha of −0.0022 is −0.22%, while a beta of 0.0059 is a 0.59 beta score). The R^2 is adjusted and describes the goodness of fit of the model.

Table 11. Robustness results of global herding models. This table shows the estimated coefficients of 2 fixed-effect herding models for Stoxx Global ESG leaders using a dataset from 1 June 2012 to 14 July 2022. Newey and West’s (1987) standard errors were used. Model I: $CSAD_t = a + \gamma_1|R_{m,t}| + \gamma_2(R_{m,t})^2 + \varepsilon_t$. Model II: $CSAD_t = a + \gamma_1D^{up}|R_{m,t}| + \gamma_2(1 - D^{up})|R_{m,t}| + \gamma_3D^{up}(R_{m,t})^2 + \gamma_4(1 - D^{up})(R_{m,t})^2 + \varepsilon_t$.

Parameter	Model I	Model II
R-squared	0.7430	0.7435
Adj. R-squared	0.7428	0.7430
Log-likelihood	11,484.76	11,486.92
a	0.0020 ***	0.0020 ***
γ1	0.6502 ***	0.6529 ***
γ2	−0.1352 *	0.6407 ***
γ3		0.2077
γ4		−0.1568
Herding (γ2 < 0)		Anti-herding (γ3 > 0, γ4 < 0)

Note: ***, * denote statistical significance at the 1% and 10% levels, respectively.

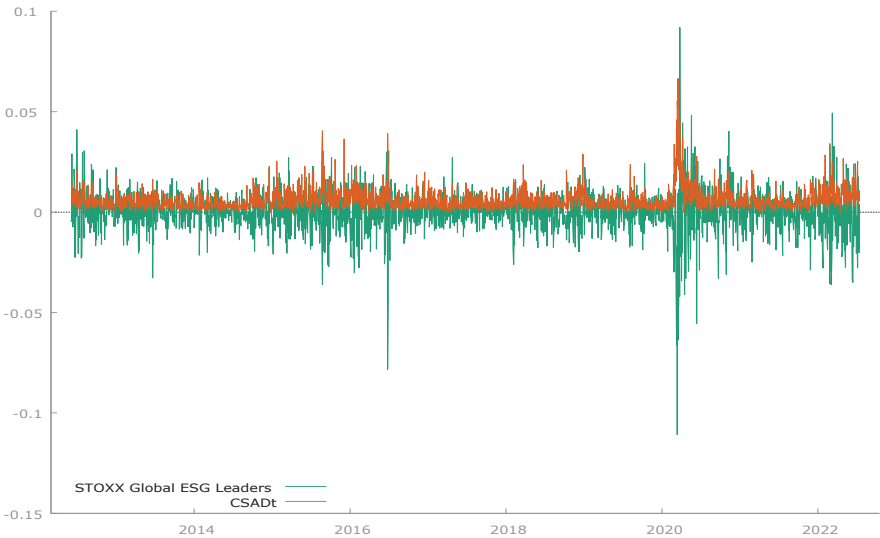


Figure 4. Herding in STOX Global ESG leaders.

5. Conclusions

ESG investing is prevalent in today’s global markets, driven by rising demand for investments that promote sustainability. Investors, fund managers, and regulators require more disclosure to evaluate the extent to which ESG leaders impact portfolio performance. We here investigated the risk-adjusted performance of ESG leaders’ portfolios in European and Global ESG leaders’ markets from 2012 to mid-2022. The ESG leaders’ portfolios did not produce significant alphas, verifying the findings of studies on negative excess returns. The regression models CAPM, FF3, and FF5 were used to test the loading factors’ validity and interpret the returns’ cross-section. We noticed sufficient evidence that the European ESG leaders tilt towards large caps, and value stocks, robust operating profitability, and low-risk investment strategies. In contrast, we reported no significant evidence in relation to Global ESG leaders regarding size, operation profitability, and investment strategy. Our results provide an effective means to capture ESG leaders’ abnormal returns by quantifying

the attribution of the ESG pillars' loading factors. Furthermore, we examined the effects and the validity of adding two secondary behavioral determinants (volatility and dispersion of returns) to the FF5 model, resulting in a novel finding of volatility tilting and a dispersion of returns bias on European and Global ESG leaders' portfolios. This finding has practical implications for investors and fund managers exposed to ESG leaders' assets in managing ESG funds and constructing sustainable portfolios. Finally, the indication of herding behavior in ESG leaders' investing eliminates the diversification benefits, leading to a risk exposure that would be difficult to hedge.

The current study contributes to the literature by providing valuable updates on factor loadings of different regression models on ESG leaders' portfolios. Furthermore, using well-known regression equation methodologies, it analyzes how those portfolios are affected by volatility, the cross-sectional dispersion of returns, and herding behavior, providing helpful insights that will help investors and policymakers to better understand pricing anomalies and behavioral finance. A limitation to be acknowledged is that the current study did not use herding estimation for sub-periods, time-varying betas of herding measures, or cross-market herding. Future research could examine how ESG investing can be applied for portfolio optimization. An interesting avenue would be to analyze and compare ESG leaders and laggards.

Author Contributions: Conceptualization, N.G. and C.F.; methodology, N.G.; software, N.G.; validation, N.G. and C.F.; formal analysis, C.F.; investigation, N.G.; resources, N.G.; data curation, N.G.; writing—original draft preparation, N.G.; writing—review and editing, C.F.; visualization, N.G.; supervision, C.F.; project administration, C.F.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are available upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Adler, Timothy, and Mark Kritzman. 2008. The cost of socially responsible investing. *The Journal of Portfolio Management* 35: 52–56. [CrossRef]
- Ameye, Nicolas, Jacques Bughin, and Nicolas van Zeebroeck. 2023. How uncertainty shapes herding in the corporate use of artificial intelligence technology. *Technovation* 127: 102846. [CrossRef]
- Arjoon, Vaalmikki, and Chandra Shekhar Bhatnagar. 2017. Dynamic herding analysis in a frontier market. *Research in International Business and Finance* 42: 496–508. [CrossRef]
- Auer, Benjamin R., and Frank Schuhmacher. 2016. Do socially (ir)responsible investments pay? New evidence from international ESG data. *The Quarterly Review of Economics and Finance* 59: 51–62. [CrossRef]
- Babalos, Vassilios, and Stavros Stavroyiannis. 2015. Herding, anti-herding behavior in metal commodities futures: A novel portfolio-based approach. *Applied Economics* 47: 4952–66.
- Bauer, Rob, Kees Koedijk, and Roger Otten. 2005. International evidence on ethical mutual fund performance and investment style. *Journal of Banking & Finance* 29: 1751–67. [CrossRef]
- Benz, Lukas, Andrea Jacob, Stefan Paulus, and Marco Wilkens. 2020. Herds on green meadows: The decarbonization of institutional portfolios. *Journal of Asset Management* 21: 13–31. [CrossRef]
- Berlinger, Edina, and Anita Lovas. 2015. *Sustainability and Growth: The Stern Review and the Model of Directed Technical Change*. IEHAS Discussion Papers No. MT-DP-2015/1. Budapest: Hungarian Academy of Sciences, Institute of Economics, Centre for Economic and Regional Studies. ISBN 978-615-5447-59-4.
- Blondel, Roxane. 2022. *Biases in Investment Decision Making: A Study on Differences in Herding Behavior between Sustainable and Traditional Investments*. Ottignies-Louvain-la-Neuve: Louvain School of Management. Available online: <http://hdl.handle.net/2078.1/thesis:33810> (accessed on 7 January 2023).
- Caoa, Ying, Anna von Reibnitz, and Geoffrey J. Warren. 2019. Return Dispersion and Fund Performance: Australia—The Land of Opportunity? Asian-Pacific Res. Exchange. Available online: <https://www.arx.cfa/-/media/regional/arx/post-pdf/2019/11/08/return-dispersion-and-fund-performance---afbc2019.pdf> (accessed on 7 January 2023).
- Carhart, Mark M. 1997. On persistence in mutual fund performance. *The Journal of Finance* 52: 57–82. [CrossRef]
- Chang, Eric C., Joseph W. Cheng, and Ajay Khorana. 2000. An Examination of Herd Behavior in Equity Markets: An International Perspective. *Journal of Banking & Finance* 24: 1651–79. [CrossRef]

- Chava, Sudheer. 2014. Environmental externalities and cost of capital. *Management Science* 60: 2223–47. Available online: <https://www.jstor.org/stable/24550583> (accessed on 2 January 2023).
- Chen, Xiaoying, and Nicholas Ray-Wang Gao. 2020. Revisiting Fama–French’s asset pricing model with an MCB volatility risk factor. *The Journal of Risk Finance* 21: 233–51. [CrossRef]
- Chiang, Shu-Mei. 2012. The relationships between implied volatility indexes and spot indexes. *Procedia—Social and Behavioral Sciences* 57: 231–35. [CrossRef]
- Chiang, Thomas C., and Dazhi Zheng. 2010. An Empirical Analysis of Herd Behavior in Global Stock Markets. *Journal of Banking & Finance* 34: 1911–21. [CrossRef]
- Chiang, Thomas Chinan, Jiandong Li, Lin Tan, and Edward Nelling. 2013. Dynamic herding behavior in Pacific-Basin markets: Evidence and implications. *Multinational Finance Journal* 17: 165–200. Available online: <https://ssrn.com/abstract=2278979> (accessed on 2 January 2023).
- Ciciretti, R., A. Dalò, and G. Ferri. 2021. Herding and Anti-Herding Across ESG Funds. Working Paper. 524. Available online: <https://ssrn.com/abstract=3957232> (accessed on 12 January 2023).
- Cocozza, Rosa, Domenico Curcio, and Antonio Pacifico. 2021. Do Global Markets Imply Common Fear? *Rivista Bancaria Minerva Bancaria*. Available online: <https://ssrn.com/abstract=3747907> (accessed on 12 January 2023).
- Consolandi, Costanza, Ameeta Jaiswal-Dale, Elisa Poggiani, and Alessandro Vercelli. 2009. Global standards and ethical stock indexes: The case of the Dow Jones Sustainability Stoxx index. *Journal of Business Ethics* 87: 185–97. [CrossRef]
- Coskun, Esra Alp, Chi Keung Marco Lau, and Hakan Kahyaoglu. 2020. Uncertainty and Herding Behavior: Evidence from Cryptocurrencies. *Research in International Business and Finance* 54: 101284. Available online: <http://www.sciencedirect.com/science/article/pii/S0275531920300957> (accessed on 7 January 2023).
- de Oliveira, Erick Meira, Felipe Arias Fogliano de Souza Cunha, Rafael Baptista Palazzi, Marcelo Cabus Klotzle, and Paula Medina Maçaira. 2020. On the effects of uncertainty measures on sustainability indices: An empirical investigation in a nonlinear framework. *International Review of Financial Analysis* 70: 101505. [CrossRef]
- Demirer, Riza, and Shrikant P. Jategaonkar. 2013. The conditional relation between dispersion and return. *Review of Financial Economics* 22: 125–34. [CrossRef]
- Dewick, Paul R., and Shuangzhe Liu. 2022. Copula Modelling to Analyse Financial Data. *Journal of Risk and Financial Management* 15: 104. [CrossRef]
- Dhasmana, Samridhdi, Sajal Ghosh, and Kakali Kanjilal. 2023. Does investor sentiment influence ESG stock performance? Evidence from India. *Journal of Behavioral and Experimental Finance* 37: 100789. [CrossRef]
- Durand, Robert B., Dominic Lim, and J. Kenton Zumwalt. 2011. Fear and the Fama-French factors. *Applied Economics* 50: 3654–63. [CrossRef]
- Economou, Fotini, Alexandros Kostakis, and Nikolaos Philippas. 2011. Cross-Country Effects in Herding Behavior: Evidence from Four South European Markets. *Journal of International Financial Markets, Institutions and Money* 21: 443–60. Available online: <https://ssrn.com/abstract=1791582> (accessed on 12 January 2023).
- Economou, Fotini, Alexandros Kostakis, and Nikolaos Philippas. 2018. Investors’ fear and herding in the stock market. *Applied Economics* 50: 3654–63. [CrossRef]
- Economou, Fotini, Konstantinos Gavrilidis, Abhinav Goyal, and Vasileios Kallinterakis. 2015. Herding dynamics in exchange groups: Evidence from Euronext. *Journal of International Financial Markets, Institutions and Money* 34: 228–44. [CrossRef]
- Erol, Isil, Umut Unal, and Yener Coskun. 2023. ESG investing and the financial performance: A panel data analysis of developed REIT markets. *Environmental Science and Pollution Research* 30: 85154–69. [CrossRef]
- Fama, Eugene F., and Kenneth R. French. 1993. Common risk factors in the return on stocks and Bonds. *Journal of Financial Economics* 33: 3–56. [CrossRef]
- Fama, Eugene F., and Kenneth R. French. 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116: 1–22. [CrossRef]
- Fu, Jingxue, and Lan Wu. 2021. Regime-switching herd behavior: Novel evidence from the Chinese A-share market. *Finance Research Letters* 39: 101652. [CrossRef]
- Gavrilakis, Nektarios, and Christos Floros. 2023a. ESG performance, herding behavior, and stock market returns evidence from Europe. *Operational Research* 23: 3. [CrossRef]
- Gavrilakis, Nektarios, and Christos Floros. 2023b. Global uncertainties, volatility return and dependence structure on thematic investing using GARCH and copula approaches. *Reference Module in Social Sciences*. [CrossRef]
- Giese, Guido, Linda-Eling Lee, Dimitris Melas, Zoltán Nagy, and Laura Nishikawa. 2019. Foundations of ESG Investing: How ESG Affects Equity Valuation, Risk, and Performance. *The Journal of Portfolio Management* 45: 69–83. Available online: <https://www.msci.com/documents/10199/03d6faef> (accessed on 7 January 2023).
- Gleason, Kimberly C., Ike Mathur, and Mark A. Peterson. 2004. Analysis of intraday herding behavior among the sector ETFs. *Journal of Empirical Finance* 11: 681–94. [CrossRef]
- Górka, Joanna, and Katarzyna Kuziak. 2021. Volatility Modeling and Dependence Structure of ESG and Conventional Investments. *Risks* 10: 20. [CrossRef]
- Guo, Bin, Wei Zhang, Yongjie Zhang, and Han Zhang. 2017. The five-factor asset pricing model tests the Chinese stock market. *Pacific-Basin Finance Journal* 43: 84–106. [CrossRef]

- Hamilton, Sally, Hoje Jo, and Meir Statman. 1993. Doing well while doing good? The investment performance of socially responsible mutual funds. *CFA Magazine* 49: 62–66. [CrossRef]
- Hartzmark, Samuel M., and Abigail B. Sussman. 2019. Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows. *The Journal of Finance* 74: 2789–837. [CrossRef]
- Heinkel, Robert, Alan Kraus, and Josef Zechner. 2001. The effect of green investment on corporate behavior. *Journal of Financial and Quantitative Analysis* 36: 431. [CrossRef]
- Henker, Julia, Thomas Henker, and Anna Mitsios. 2004. Do Investors Herd Intraday in Australian Equities? *International Journal of Managerial Finance* 2: 196–219. [CrossRef]
- Henriksson, Roy, Joshua Livnat, Patrick Pfeifer, and Margaret Stumpp. 2019. Integrating ESG in portfolio construction. *The Journal of Portfolio Management* 45: 67–81. [CrossRef]
- Hwang, Soosung, and Mark Salmon. 2004. Market stress and herding. *Journal of Empirical Finance* 11: 585–616. [CrossRef]
- Jain, Mansi, Gagan Deep Sharma, and Mrinalini Srivastava. 2020. Can Sustainable Investment Yield Better Financial Returns: A Comparative Study of ESG and MSCI Indices. *Risks* 7: 15. [CrossRef]
- Jiang, Xiaoquan. 2010. Return dispersion and expected returns. *Financial Markets and Portfolio Management* 24: 107–35. [CrossRef]
- Kenneth R. French Library. 2023. Available online: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html (accessed on 7 January 2023).
- López-Cabarcos, M. Ángeles, Ada M. Pérez-Pico, and M. Luisa López-Pérez. 2019. Does social network sentiment influence the S&P500 environmental and socially responsible index? *Sustainability* 11: 320. Available online: <https://www.mdpi.com/2071-1050/11/2/32000> (accessed on 12 January 2023).
- Luo, Di. 2022. ESG, liquidity, and stock returns. *Journal of International Financial Markets, Institutions and Money* 78: 101526. [CrossRef]
- Managi, Shunsuke, Tatsuyoshi Okimoto, and Akimi Matsuda. 2012. Do socially responsible investment indexes outperform conventional indexes? *Applied Financial Economics* 22: 1511–27. [CrossRef]
- Marquit, M., and B. Curry. 2023. The VIX Volatility Index. *Forbes*. Available online: <https://www.forbes.com/advisor/investing/vix-volatility-index/> (accessed on 12 January 2023).
- McKenzie, Michael, and Heather Mitchell. 2002. Generalized asymmetric power ARCH modeling of exchange rate volatility. *Applied Financial Economics* 12: 555–64. [CrossRef]
- Merton, Robert C. 1987. A simple model of capital market equilibrium with incomplete information. *The Journal of Finance* 42: 483–510. [CrossRef]
- Mollet, Janick Mollet, and Andreas Ziegler. 2014. Socially responsible investing and stock performance: New empirical. *Review of Financial Economics* 23: 208–16. [CrossRef]
- Morales, Lucía, Amparo Soler-Domínguez, and James Hanly. 2019. The power of ethical investment in the context of political uncertainty. *Journal of Applied Economics* 22: 554–80. [CrossRef]
- Morningstar Manager Research. 2023. Global Sustainable Fund Flows: Q2 2023 in Review. Available online: <https://www.morningstar.com/lp/global-esg-flows> (accessed on 7 January 2023).
- Morningstar/Sustainalytics. 2023. Available online: <https://www.sustainalytics.com/esg-data> (accessed on 12 January 2023).
- Naffa, Helena, and Máté Fain. 2022. A factor approach to the performance of ESG leaders and laggards. *Finance Research Letters* 44: 102073. [CrossRef]
- Nagler, Thomas, Daniel Krüger, and Aleksey Min. 2022. Stationary vine copula models for multivariate time series. *Journal of Econometrics* 227: 305–24. [CrossRef]
- Newey, Whitney K., and Kenneth D. West. 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55: 703–8. [CrossRef]
- Nishant, S., S. Maffina, and S. Borikar. 2022. Exploring the Link Between Stock Price Volatility and ESG Scores. Insights ESG. State Street Global Advisors. Available online: <https://www.ssga.com/library-content/pdfs/ic/link-between-volatility-and-esg-scores.pdf> (accessed on 28 January 2023).
- Öcal, Hüseyin, and Anton Abdulbasah Kamil. 2021. The impact of risk indicators on sustainability (ESG) and broad-based indices: An empirical analysis from Germany, France, Indonesia and Turkey. *International Journal of Sustainable Economy* 13: 18. [CrossRef]
- Ouchen, Abdessamad. 2022. Is the ESG portfolio less turbulent than a market benchmark portfolio? *Risk Management* 24: 1–33. [CrossRef]
- Patton, Andrew J. 2012. A review of copula models for economic time series. *Journal of Multivariate Analysis* 110: 4–18. [CrossRef]
- Pedersen, Lasse Heje, Shaun Fitzgibbons, and Lukasz Pomorski. 2021. Responsible investing: The ESG-efficient frontier. *Journal of Financial Economics* 142: 572–97. [CrossRef]
- Philippas, Nikolaos, Fotini Economou, Vassilios Babalos, and Alexandros Kostakis. 2013. Herding Behavior in REITs: Novel Tests and the Role of Financial Crisis. *International Review of Financial Analysis* 29: 166–74. [CrossRef]
- PricewaterhouseCoopers. 2022. The Growth Opportunity of the Century. Available online: <https://www.pwc.lu/en/sustainable-finance/esg-report-the-growth-opportunity-of-the-century.html> (accessed on 28 January 2023).
- Przychodzen, Justyna, Fernando Gómez-Bezares, Wojciech Przychodzen, and Mikel Larreina. 2016. ESG Issues among Fund Managers—Factors and Motives. *Sustainability* 8: 1078. [CrossRef]
- Qontigo. 2023. STOXX European/Global ESG Leaders' Indices. Available online: <https://qontigo.com/> (accessed on 15 January 2023).

- Ramadan, Imad Zeyad. 2015. Cross-Sectional Absolute Deviation Approach for Testing the Herd Behavior Theory: The Case of the ASE Index. *International Journal of Economics and Finance* 7: 188–93. [CrossRef]
- Renneboog, Luc, Jenke Ter Horst, and Chendi Zhang. 2008. Socially responsible investments: Institutional aspects, performance, and investor behavior. *Journal of Banking & Finance* 32: 1723–42. [CrossRef]
- Rompotis, Gerasimos G. 2018. Herding Behavior among Exchange-Traded Funds. *Journal of Behavioral Finance* 19: 483–49. [CrossRef]
- Ross, A. 2023. ESG Investors Play the Long Game. *Financial Times*. Available online: <https://www.ft.com/content/959dfa93-e6de-441e-9328-9a99c11bd54d> (accessed on 15 January 2023).
- Rubbaniy, Ghulame, Shoaib Ali, Costas Siriopoulos, and Aristeidis Samitas. 2021. Global Financial Crisis, COVID-19, Lockdown, and Herd Behavior in the US ESG Leader Stocks. Available online: <https://ssrn.com/abstract=3868114> (accessed on 8 January 2023).
- Sabbaghi, Omid. 2023. ESG and volatility risk: International evidence. *Business Ethics, the Environment & Responsibility* 32: 802–18. [CrossRef]
- Shaikh, Imlak. 2022. On the relationship between policy uncertainty and sustainable investing. *Journal of Modelling in Management* 17: 1504–23. [CrossRef]
- Siriopoulos, Costas, and Athanasios Fassas. 2009. Implied Volatility Indices—A Review. Available online: <https://ssrn.com/abstract=1421202> (accessed on 8 January 2023).
- Stanescu, Silvia, and Radu Tunaru. 2013. Investment Strategies with VIX and VSTOXX Futures. Available online: <https://ssrn.com/abstract=2351427> (accessed on 15 January 2023).
- Tao, Chen, Dong Hui, and Lin Chen. 2020. Institutional shareholders and corporate social responsibility. *Journal of Financial Economics* 135: 483–504. [CrossRef]
- Teti, Emanuele, Maurizio Dallochio, and Giulio L'Erario. 2023. The impact of ESG tilting on the performance of stock portfolios in times of crisis. *Finance Research Letters* 52: 103522. [CrossRef]
- Treynor, Jack L. 1961. Market Value, Time, and Risk. Available online: <https://ssrn.com/abstract=2600356> (accessed on 17 January 2023).
- Treynor, Jack L. 1962. Toward a Theory of Market Value of Risky Assets. Available online: <https://ssrn.com/abstract=628187> (accessed on 28 January 2023).
- Vergili, Gizem, and Mehmet Sinan Celik. 2023. The relationship between volatility indices (VIX) and sustainability (DJSIEMUP): An ARDL approach. *Business and Economics Research Journal* 14: 19–29. [CrossRef]
- Verousis, Thanos, and Nikolaos Voukelatos. 2018. Cross-Sectional Dispersion and Expected Returns. *Quantitative Finance* 18: 813–26. [CrossRef]
- Whaley, Robert E. 2000. The Investor Fear Gauge. *The Journal of Portfolio Management* 26: 12–17. [CrossRef]
- Yang, Wan-Ru, and Ming-Che Chuang. 2023. Do investors herd in a volatile market? Evidence of dynamic herding in Taiwan, China, and US stock markets. *Finance Research Letters* 52: 103364. [CrossRef]
- Yarovaya, Larisa, Roman Matkovskyy, and Akanksha Jalan. 2021. The effects of a “black swan” event (COVID-19) on herding behavior in cryptocurrency markets. *Journal of International Financial Markets, Institutions and Money* 75: 101321. [CrossRef]
- Yu, Huaibing. 2022. Does sustainable competitive advantage make a difference in stock performance during the COVID-19 pandemic? *Finance Research Letter* 48: 102893. [CrossRef] [PubMed]
- Zaremba, Adam, and Anna Czapkiewicz. 2017. Digesting anomalies in emerging European markets: A comparison of factor pricing models. *Emerging Markets Review* 31: 1–15. [CrossRef]
- Zehir, Emre, and Aslı Aybars. 2020. Is there any effect of ESG scores on portfolio performance? Evidence from Europe and Turkey. *Journal of Capital Markets Studies* 4: 129–43. [CrossRef]
- Zerbib, Olivier D. 2020. A sustainable Capital Asset Pricing Model (S-CAPM): Evidence from Green Investing and Sin Stock Exclusion. Proceedings of Paris December 2020 Finance Meeting EUROFIDAI—ESSEC. Available online: <https://ssrn.com/abstract=3455090> (accessed on 12 January 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

Option Pricing with the Logistic Return Distribution

Haim Levy * and Moshe Levy

The Jerusalem School of Business Administration, The Hebrew University of Jerusalem, Jerusalem 91905, Israel; mslm@huji.ac.il

* Correspondence: mshlevy@huji.ac.il

Abstract: The Black–Scholes model and many of its extensions imply a log-normal distribution of stock total returns over any finite holding period. However, for a holding period of up to one year, empirical stock return distributions (both conditional and unconditional) are not log-normal, but rather much closer to the logistic distribution. This paper derives analytic option pricing formulas for an underlying asset with a logistic return distribution. These formulas are simple and elegant and employ exactly the same parameters as B&S. The logistic option pricing formula fits empirical option prices much better than B&S, providing explanatory power comparable to much more complex models with a larger number of parameters.

Keywords: option pricing; distribution of returns; logistic distribution; holding period

JEL Classification: G13

1. Introduction

The seminal works of Black and Scholes (1973) and Merton (1973) are the foundation for a wide spectrum of option pricing models, incorporating stochastic volatility (see for example, Hull and White (1987); Scott (1987); Wiggins (1987); Melino and Turnbull (1990); Stein and Stein (1991); Heston (1993); Naik (1993); Duffie et al. (2000); and Barone-Adesi et al. (2008)), jumps (Merton (1976); Bates (1991) and Madan et al. (1998); Kou (2002); Eraker (2004); Cremers et al. (2008); Akerer and Filipović (2020)), stochastic interest rates (Merton (1973) and Amin and Jarrow (1992)), leverage effects (Geske (1979)), and various combinations of these factors (Bailey and Stulz (1989); Amin and Ng (1993); Bakshi et al. (1997); Pan (2002); Willems (2019); Kirkby and Nguyen (2020); Bartl et al. (2020)). In the basic B&S model, the volatility is constant, and the underlying asset price follows a geometric Brownian motion, implying that the total return distribution ($1 + \text{rate of return}$) is log-normal for any finite holding period. In the stochastic volatility models, the unconditional return distribution is a mixture of the conditional distributions across states and may thus have fat tails. However, the conditional return distribution for a given volatility level is still log-normal. Jumps may induce deviations from the log-normal distribution; however, as these jumps are typically infrequent, the bulk of the distribution remains approximately log-normal.¹

This feature of the main option pricing models is in sharp contrast to the empirical evidence: statistical tests clearly reject the log-normality of returns. For horizons of up to a year, which is the relevant horizon for most options, the empirical total return distribution is very different from the log-normal. Rather, it is much closer to the logistic distribution (Mantegna and Stanley (1995); Levy and Duchin (2004)). This is true for both the unconditional and the conditional distributions. Figures 1 and 2 illustrate this point. Figure 1 shows the empirical distribution of one-day total returns on the S&P 500 index, the best log-normal fit and the best logistic fit to this distribution. It is evident that the logistic distribution fits the data much better, and this is confirmed by formal tests. In fact, out of 10 theoretical distributions widely used in the literature, the logistic distribution provides

Citation: Levy, Haim, and Moshe Levy. 2024. Option Pricing with the Logistic Return Distribution. *Journal of Risk and Financial Management* 17: 67. <https://doi.org/10.3390/jrfm17020067>

Academic Editor: Thanasis Stengos

Received: 3 January 2024

Revised: 8 February 2024

Accepted: 8 February 2024

Published: 10 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

the best fit (we expand on these points in the next section). Figure 2 shows the same result for the return distribution conditioned on the previous day’s VIX. This empirical observation of the rejection of log-normality in favor of the logistic distribution is the motivation for the present paper.

This paper analytically derives option pricing formulas for the case of the logistic return distribution. Our approach is different than the standard option pricing approach in the following sense: rather than starting from an instantaneous return process, our starting point is the distribution of the underlying asset’s value at expiration. We do not make any assumptions about the instantaneous return process (such as a Brownian motion). Instead, we directly assume a logistic distribution of stock prices at expiration, and we employ the no-arbitrage argument to derive the call option price via risk-neutral valuation. From this price we also derive the hedge ratio facilitating continuous hedging. What is the origin of the empirically observed logistic return distribution? We suggest a possible explanation, but this is rather speculative, and alternative explanations are certainly possible. The important point is that regardless of the reason for the logistic stock price distribution at expiration, this distribution allows us to derive a simple analytical formula for option prices.

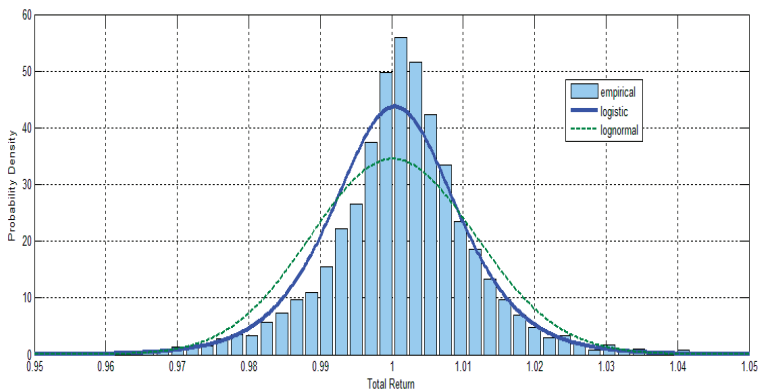


Figure 1. The unconditional daily return distribution for the S&P 500 index. The dashed line is the best log-normal fit. The bold line is the best logistic fit. The logistic distribution provides the best fit of all 10 theoretical distributions examined. The description of the data and more details are provided in Section 2.

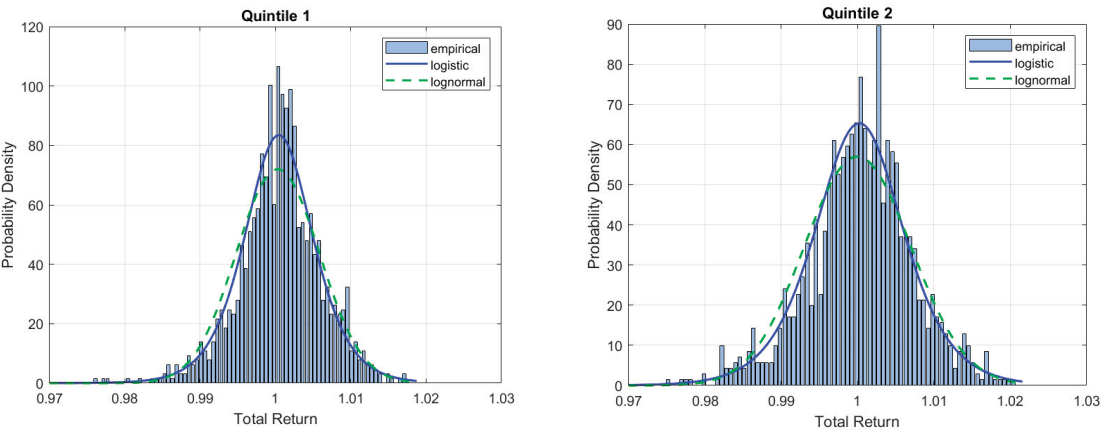


Figure 2. Cont.

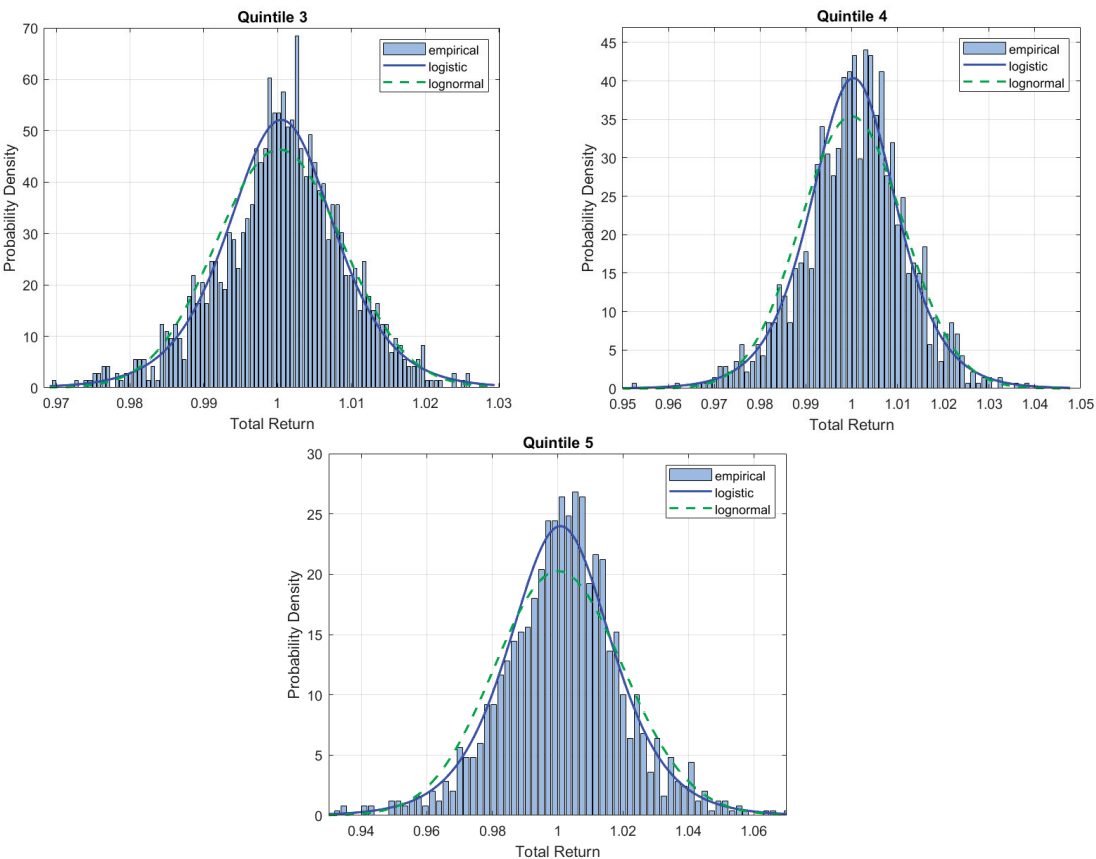


Figure 2. The distributions of daily total returns conditioned on the pervious-day’s VIX. Quintile 1 is the one with the lowest VIX. For all five quintiles, the logistic distribution fits better than the log-normal (see also Tables 1 and 3).

Table 1. Goodness of fit for the conditional and unconditional daily return distribution.

	Unconditional	Quintile 1 (Lowest VIX)	Quintile 2	Quintile 3	Quintile 4	Quintile 5 (Highest VIX)
Logistic	−18,958	−4566	−4274	−4010	−3701	−3063
Log-logistic	−18,955	−4565	−4273	−4009	−3700	−3062
Rice	−18,399	−4534	−4253	−4005	−3683	−3012
Normal	−18,399	−4534	−4253	−4005	−3683	−3011
Gamma	−18,396	−4533	−4252	−4004	−3681	−3011
Log-normal	−18,394	−4532	−4251	−4004	−3680	−3010

This unorthodox approach has pros and cons relative to the traditional approach of starting with an instantaneous return process. The disadvantages are that (1) the model does not explain *why* the distribution is logistic and (2) the logistic distribution can only serve as an approximation because it is not limited to the positive domain; in addition, the sum of two logistic random variables is not exactly logistically distributed (thus, if the one-day return distribution is exactly logistic, the two-day return distribution can only

be approximately logistic). We advocate that these limitations are not severe, because even if the logistic distribution only approximates the true distribution, for the relevant holding periods this approximation is very good (and much better than the log-normal), and it is thus very useful. The advantage of our approach is that it yields a simple analytic option pricing formula that is based directly on the empirically observed return distribution. We believe that the non-standard approach suggested here complements the traditional approach and yields a new and useful perspective for option pricing.

The option pricing formulas derived for the logistic case are surprisingly simple and elegant. They employ the same parameters as B&S, with a single variable that needs to be estimated: volatility. They are even easier to implement than B&S. Thus, the challenging econometric issues involved with the sophisticated option pricing models are circumvented. Empirically, the logistic option pricing formulas perform much better than B&S and are comparable to more complex option pricing models with a large number of parameters. We view this model as a first step. One could potentially extend the logistic option pricing model presented here to incorporate stochastic volatility, stochastic interest rates, etc., which may further improve the performance of the model. Of course, this would come at a price of increased complexity. For the sake of clarity, in this paper we present the simplest logistic option pricing model, which can serve as a springboard for further extensions.

The structure of this paper is as follows. The next section shows that the empirical return distribution is best fitted by the logistic distribution, which is the motivation of the paper. Section 3, which is the main contribution of the paper, derives option prices under the logistic return distribution. Section 4 examines the empirical performance of the logistic option pricing formula. In Section 5, we suggest a possible explanation for the empirically observed logistic return distribution. Section 6 concludes.

2. The Empirical Return Distribution

It has been long known that the log-normal distribution does not fit the empirical return distribution well. Focusing on short-horizon returns, Mandelbrot (1963) and Fama (1965) suggest that the stable Pareto (Lévy) distribution provides a better fit. Officer (1972) suggests that the distribution should have fat tails but a finite second moment. These ideas are supported by Mantegna and Stanley (1995), who find that a truncated stable Paretian distribution fits the empirical return distribution well for holding periods of up to about a month (more on the relation between the truncated stable Paretian distribution and the logistic distribution in Section 5). One of the most comprehensive studies comparing the empirical return distribution with a large number of theoretical distributions is that of Levy and Duchin (2004). They examine the fit of many different distributions as a function of the holding period and conclude that:

“For investment horizons shorter than one year the logistic distribution best describes the empirical data” (p. 61).

They find this result for stock indices, individual stocks, and even bonds.²

In examining the shape of the return distribution, we confirm and extend the results of Levy and Duchin (2004) in the following ways. First, we employ a different and longer sample period. Second, we include in our analysis several additional theoretical candidate distributions not considered in the Levy and Duchin analysis. Third, we employ several different criteria for goodness-of-fit. Most importantly, we examine not only the unconditional return distribution, but also the return distribution conditional on the anticipated volatility, as captured by the VIX index.

We examine the returns of the S&P 500 index for holding periods of 1 day up to 1 year, which is the range of horizons relevant for most options. For the one-day horizon we employ the daily returns on the S&P 500 index over the period from January 1990–December 2013, as reported in the CRSP files. We test the goodness-of-fit to the empirical return distribution by 10 theoretical distributions employed in the literature: log-normal, normal, logistic, gamma, inverse-Gaussian, log-logistic, Weibull, Rice, generalized extreme value, and beta

distributions. We employ two criteria for goodness-of-fit: negative log-likelihood and the Bayesian information criterion.

For a general impression, Figure 1 plots the empirical daily return histogram, the best log-normal fit, and the best logistic fit. The logistic distribution clearly fits better than the log-normal. This is confirmed by the formal tests. When comparing the 10 potential distributions with the empirical data, we find that the distribution that provides the best fit is the logistic distribution, followed by the log-logistic distribution, the Rice distribution, the normal distribution, and the gamma distribution. The log-normal distribution comes in sixth place (see Table 1). We obtain the exact same ranking with the three different goodness-of-fit criteria.

As the market is characterized by periods of high daily volatility and other periods of low daily volatility, one could argue that the unconditional return distribution is a mixture of several different distributions corresponding to different market conditions. In order to examine the conditional return distribution, we partition the return sample into five quintiles, sorted by the previous day’s level of the VIX index. For each quintile, the conditional return distribution is the set of next-day returns. Hence, we have five different empirical conditional return distributions. For each of these distributions, we examine which theoretical distribution fits best. Table 2 provides some summary statistics on the VIX quintiles and the five conditional return distributions. As expected, the variance in the return distribution increases with the VIX quintile, implying that there is a positive relation between the VIX and the variance in the next-day return. Figure 2 shows the conditional daily return distributions for the five quintiles, as well as the best log-normal fit and best logistic fit for each distribution. The figure reveals that the logistic distribution provides a better fit than the log-normal for each of the five conditional distributions. This is confirmed by the formal goodness-of-fit test: the logistic distribution provides the best fit for each of the five conditional empirical distributions, followed by the log-logistic and the Rice distributions. We obtain the same ranking for all five conditional distributions, and by all three goodness-of-fit criteria. Table 1 reports the negative log-likelihood for all five conditional return distributions, as well as for the unconditional return distribution (recall that the more negative the statistic, the better the fit).

Table 2. Summary statistics.

Quintile	1	2	3	4	5
VIX range (%)	9.3–13.6	13.6–16.9	16.9–20.5	20.5–25.1	25.1–80.9
Mean of next-day return: (%)	0.04	0.01	0.01	0.02	0.07
Standard deviation of next-day return: (%)	0.56	0.70	0.86	1.14	1.97

The table reports the negative log-likelihood for various candidate distributions to fit the empirical daily return distribution. The first column corresponds to the unconditional return distribution, as shown in Figure 1. The next columns correspond to the daily return distribution conditional on the previous day’s VIX, as reported in Figure 2. In all cases the logistic distribution provides the best fit, while the lognormal comes in sixth place (recall that the more negative the negative log-likelihood, the better the fit). The exact same ranking is obtained with the BIC goodness-of-fit criterion.

How does the length of holding period affect the return distribution? In order to examine this issue, we analyze the one-month, three-month, six-month, nine-month, and twelve-month return distributions of the S&P 500 index. For each holding period we report the six distributions that provide the best fit, in order of their performance, by the two goodness-of-fit criteria. The results are shown in Table 3. As the table reveals, the logistic distribution provides the best fit for all horizons shorter than a year. This holds for both of the criteria employed. Raftery (1995) interprets a difference of 0–2 in the BIC as “weak” evidence, a difference of 2–6 as “positive” evidence, 6–10 as “strong” evidence, and a BIC difference exceeding 10 as “very strong” evidence. Table 3 reveals that for monthly

returns, the evidence in favor of the logistic distribution is “very strong” in comparison to all other alternatives. As the horizon increases, the difference between the logistic and the log-logistic, which provides the next best fit, decreases. For annual horizons the distribution that fits best is either the Weibull distribution or the generalized extreme value distribution, depending on the criterion employed. This is consistent with the findings of Levy and Duchin, who find that the logistic distribution dominates for horizons of up to a year but the Weibull distribution provides the best fit for annual returns (see Levy and Duchin 2004, Exhibit 3 on p. 53).³

The 1990–2013 daily sample is partitioned into five quintiles, sorted by the VIX. We create five conditional distributions of the next-day returns. For example, the return distribution of quintile 1 is the set of returns on the days following the days with the lowest VIX, etc. The second row shows the VIX cut-off points for each quintile, and the third and fourth rows provide the mean return and standard deviation of returns for the associated conditional return distributions. As expected, higher VIX values are associated with more volatile next-day returns. The entire distributions are shown in Figure 2.

For each holding-period, the table ranks the six distributions that provide the best fit to the empirical returns, in order of their performance. The logistic distribution provides the best fit for all horizons shorter than a year, by both log-likelihood and the BIC criterion.

Table 3. The return horizon and the best-fitting distributions.

Goodness-of-Fit Criterion	1 Month	3 Months	6 Month	9 Months	12 Months
Negative log-likelihood	logistic—1667	logistic—324	logistic—101	logistic—42	general extreme value—22
	log-logistic—1660	log-logistic—322	log-logistic—97	Weibull—41	Weibull—21
	normal—1568	log-normal—286	normal—94	normal—40	normal—19
	Rice—1568	inverse	Rice—94	Rice—40	Rice—19
	Gamma—1566	Guassian—285	gamma—90	general extreme value—39	logistic—19
Bayesian information criterion	log-normal—1562	gamma—285	log-normal—86	log-logistic—37	gamma—16
	logistic—3319	logistic—635	logistic—192	logistic—73	Weibull—34
	log-logistic—3306	log-logistic—632	log-logistic—185	Weibull—72	normal—30
	normal—3122	log-normal—561	normal—177	normal—70	Rice—30
	Rice—3122	inverse	Rice—177	Rice—70	general extreme value—30
	gamma—3118	Guassian—559	gamma—170	general extreme value—64	logistic—28
	log-normal—3110	gamma—558	log-normal—162	log-logistic—64	gamma—23
		Rice—537			

Most options have maturities less than a year. The relevant return distribution for pricing these options is the logistic distribution. For horizons of up to 6 months, the distribution which provides the second-best fit is the log-logistic. This distribution has the theoretical advantage of total returns being restricted to the positive domain. However, even for annual returns, a negative total return represents an event that is about 5.5 standard deviations to the left of the mean, so this is not a major consideration. The logistic return distribution not only provides the best fit to the empirical distribution, but it also implies very elegant option pricing formulas, as shown in the next section.

3. Logistic Option Pricing

Consider an underlying asset with a value at maturity distributed according to some probability density function $f(x)$. The absence of arbitrage implies that one can employ risk-neutral valuation to price options on this asset. In a risk-neutral world, the value of a European Call option with strike price K and maturity T is given by:

$$C_0 = \frac{1}{(1+r_f)^T} \int_K^\infty (x-K)f(x)dx, \quad (1)$$

where C_0 is the Call price and r_f is the risk-free rate. In the risk-neutral world, the current price of the underlying asset is its expected value at maturity, discounted at the risk-free rate: $S_0 = \frac{1}{(1+r_f)^T} E[x]$. One way to obtain the B&S formula is to plug the log-normal distribution of terminal asset values implied by the B&S geometric Brownian motion as $f(x)$ in Equation (1) (see, for example, Cox et al. (1979); and Hull (2009))⁴. In what follows, we employ Equation (1) with the logistic distribution for $f(x)$ to obtain logistic option pricing.

The logistic distribution p.d.f. is given by⁵:

$$f(x; \mu, s) = \frac{e^{-\frac{x-\mu}{s}}}{s \left(1 + e^{-\frac{x-\mu}{s}}\right)^2}, \quad (2)$$

where μ is the mean of the distribution and the standard deviation is $\sigma = \frac{\pi s}{\sqrt{3}}$. It is convenient to express the logistic distribution and its integrals using the hyperbolic functions defined as (see, for example, Becker 1931):

$$\sinh(x) \equiv \frac{e^x - e^{-x}}{2}, \cosh(x) \equiv \frac{e^x + e^{-x}}{2}, \tanh(x) \equiv \frac{e^x - e^{-x}}{e^x + e^{-x}}, \operatorname{sech}(x) \equiv \frac{2}{e^x + e^{-x}}. \quad (3)$$

The logistic distribution (2) can be equivalently written as:

$$f(x; \mu, s) = \frac{1}{4s} \operatorname{sech}^2\left(\frac{x-\mu}{2s}\right). \quad (4)$$

Plugging this distribution as $f(x)$ in Equation (1), we obtain:

$$C_0 = \frac{1}{(1+r_f)^T} \int_K^\infty (x-K) \frac{1}{4s} \operatorname{sech}^2\left(\frac{x-\mu}{2s}\right) dx. \quad (5)$$

Employing the following change in variables: $z \equiv \frac{x-\mu}{2s}$ $dz = \frac{1}{2s} dx$ yields:

$$C_0 = \frac{1}{(1+r_f)^T} \int_{\frac{K-\mu}{2s}}^\infty (2sz + \mu - K) \frac{1}{4s} \operatorname{sech}^2(z) 2s dz,$$

or:

$$C_0 = \frac{1}{(1+r_f)^T} \int_{\frac{K-\mu}{2s}}^\infty \left(s z + \frac{\mu-K}{2}\right) \operatorname{sech}^2(z) dz \quad (6)$$

Making use of the relations:

$(\tanh(x))' = \operatorname{sech}^2(x)$ and $\int x \operatorname{sech}^2(x) dx = x \cdot \tanh(x) - \ln(\cosh(x))$ leads to:

$$\begin{aligned} C_0 &= \frac{1}{(1+r_f)^T} \left[\frac{\mu-K}{2} \tanh(z) + s z \tanh(z) - s \cdot \ln(\cosh(z)) \right]_{\frac{K-\mu}{2s}}^\infty = \\ C_0 &= \frac{1}{(1+r_f)^T} \left\{ \left[\frac{\mu-K}{2} + s z - s z + s \cdot \ln(2) \right] - \left[\frac{\mu-K}{2} \tanh\left(\frac{K-\mu}{2s}\right) + \frac{K-\mu}{2} \tanh\left(\frac{K-\mu}{2s}\right) - s \cdot \ln\left(\cosh\left(\frac{K-\mu}{2s}\right)\right) \right] \right\} \quad (7) \\ C_0 &= \frac{1}{(1+r_f)^T} \left[\frac{\mu-K}{2} + s \cdot \ln(2) + s \cdot \ln\left(\cosh\left(\frac{K-\mu}{2s}\right)\right) \right], \end{aligned}$$

or:

$$C_0 = \frac{1}{(1+r_f)^T} \left[\frac{\mu - K}{2} + s \cdot \ln \left(2 \cosh \left(\frac{K - \mu}{2s} \right) \right) \right]. \quad (8)$$

As in B&S, under risk-neutrality the underlying asset's expected value grows at the risk-free rate. Thus, if we denote the underlying asset's current price with S_0 , we obtain: $\mu = S_0(1+r_f)^T$ (recall that μ is the expected value of the end-of-period asset value). $\frac{\pi s}{\sqrt{3}}$ is the standard deviation of the end-of-period underlying asset value. We can express this instead as the standard deviation of returns from date 0 to maturity: $\sigma = \frac{1}{S_0} \frac{\pi s}{\sqrt{3}}$, or: $s = \sigma S_0 \frac{\sqrt{3}}{\pi}$. Plugging these values of μ and s in Equation (8), we finally obtain:

$$C_0 = \frac{S_0}{2} - \frac{K}{2(1+r_f)^T} + \frac{1}{(1+r_f)^T} \frac{\sqrt{3}}{\pi} S_0 \sigma \cdot \ln \left(2 \cosh \left(\frac{(K/S_0) - (1+r_f)^T}{2 \frac{\sqrt{3}}{\pi} \sigma} \right) \right). \quad (9)$$

This is the option pricing formula under the logistic return distribution. It is straightforward to verify that $C_0 \rightarrow S_0$ as $K \rightarrow 0$ and that $C_0 \rightarrow 0$ as $K \rightarrow \infty$, as in B&S.

The logistic option pricing formula is based on exactly the same parameters as the B&S formula, and it is very easy to implement. Just as in B&S, there is a single parameter that needs to be estimated—the standard deviation. Note, however, that the notation for the standard deviation is slightly different here and in B&S. In B&S, σ denotes the annual standard deviation of log-returns; thus, the standard deviation of returns from date 0 until expiration is $\sigma\sqrt{T}$. Here, σ simply denotes the standard deviation of returns from date 0 until expiration.⁶

To get a feeling for the logistic option pricing relative to the B&S pricing, consider an underlying asset with $S_0 = \$100$ and $\sigma = 20\%$, a risk-free rate of 1%, and maturity in 1 year. For an in-the-money option with $K = \$80$, the B&S Call price is USD 26.22. The logistic option price for this option, given by Equation (9), is USD 26.73. Thus, in this case the logistic price is 1.95% higher than the B&S price. For the at-the-money option with $K = \$100$, the B&S price is USD 8.28 while the logistic price is USD 8.07. In this case, the logistic price is 2.5% lower than the B&S price. Finally, if the option is out-of-the-money with $K = \$120$, the B&S price is USD 2.22 while the logistic price is USD 1.79. In this case, the relative pricing difference is quite large, with the logistic option priced 19.4% below the B&S price.

We can derive the price of a European Put by employing Put–Call parity. This yields:

$$P_0 = C_0 - S_0 + \frac{K}{(1+r_f)^T} \Rightarrow$$

$$P_0 = -\frac{S_0}{2} + \frac{K}{2(1+r_f)^T} + \frac{1}{(1+r_f)^T} \frac{\sqrt{3}}{\pi} S_0 \sigma \cdot \ln \left(2 \cosh \left(\frac{(K/S_0) - (1+r_f)^T}{2 \frac{\sqrt{3}}{\pi} \sigma} \right) \right). \quad (10)$$

The formula for the price of the Put is exactly the same as the one for the Call, only with the signs of the first two terms reversed.

It is straightforward to calculate the hedge ratio Δ from Equation (9):

$$\Delta \equiv \frac{\partial C_0}{\partial S_0} = \frac{1}{2} \left[1 - \tanh \left(\frac{(K/S_0) - (1+r_f)^T}{2 \frac{\sqrt{3}}{\pi} \sigma} \right) \right]. \quad (11)$$

Γ , the rate of change in Δ with respect to the underlying asset's price, is given by:

$$\Gamma \equiv \frac{\partial^2 C_0}{\partial S_0^2} = \frac{\pi(1+r_f)^T}{4\sqrt{3}\sigma S_0} \cdot \sec h^2 \left(\frac{(K/S_0) - (1+r_f)^T}{2\frac{\sqrt{3}}{\pi}\sigma} \right). \quad (12)$$

V, the rate of change in the option value with respect to the underlying asset's volatility is given by:

$$V = \frac{1}{(1+r_f)^T} \frac{\sqrt{3}}{\pi} \left[\ln \left(2 \cosh \left(\frac{(K/S_0) - (1+r_f)^T}{2\frac{\sqrt{3}}{\pi}\sigma} \right) \right) - \frac{1}{\sigma S_0} \left(\frac{(K/S_0) - (1+r_f)^T}{2\frac{\sqrt{3}}{\pi}\sigma} \right) \tanh \left(\frac{(K/S_0) - (1+r_f)^T}{2\frac{\sqrt{3}}{\pi}\sigma} \right) \right]. \quad (13)$$

Just as in B&S, pricing and hedging require the estimation of only the volatility.

4. Empirical Performance of the Logistic Option Pricing Formula

In this section, we compare the empirical performance of the logistic option pricing formula with B&S and more complex option pricing models along two dimensions: the implied volatility smile and dynamic hedging errors.

4.1. Data

We employ daily prices of all call options on the S&P 500 index in the OptionMetrics file in the period 1996–2013. To exclude low-liquidity options, with notoriously unreliable price records, we impose the following three requirements:

- The bid-ask spread is less than 2.5% of (bid + ask)/2.
- Bid > USD 10
- Open interest > 500

With this screening, we have a total of 61,820 option price observations. The risk-free rate is taken as the corresponding T-Bill rate. To match the rate until expiration, we extrapolate linearly between the two T-Bills with the closest maturities. Following Hull (2009), the underlying asset price is taken as the level of the S&P 500 index minus the present value of all dividends until expiration.

4.2. Volatility Smile

For each option price, we numerically calculate the implied B&S volatility and the volatility implied by the logistic option price formula, Equation (9). Figure 3 shows the implied volatilities as a function of moneyness for the two models. To get more detailed results, we partition the sample according to days to expiration. Panel A corresponds to options with 25–50 days to expiration, Panel B to options with 50–75 days to expiration, and Panel C to options with 75–100 days to expiration.

Note that if the model fits the data well, no smile is expected. While the smile is not completely eliminated with the logistic formula, it is greatly attenuated relative to the one obtained with B&S. Notice that moneyness range in Figure 3 is much wider than the range usually reported in the literature. In the more typically employed narrower moneyness range, the variability in the implied volatility under the logistic model is comparable to that obtained with more complex models with a larger number of parameters. For example, Bakshi et al. (1997, Figure 1, p. 2022) report that the implied volatility for various stochastic volatility models (with and without jumps) and 60–180 days to expiration varies by about 3%–4% in the $0.93 < S/K < 1.08$ range. We find similar variability in the implied volatility for the logistic pricing formula over the same moneyness range and similar maturities. This similar performance is obtained with the very simple logistic option pricing model, which has only a single free parameter.

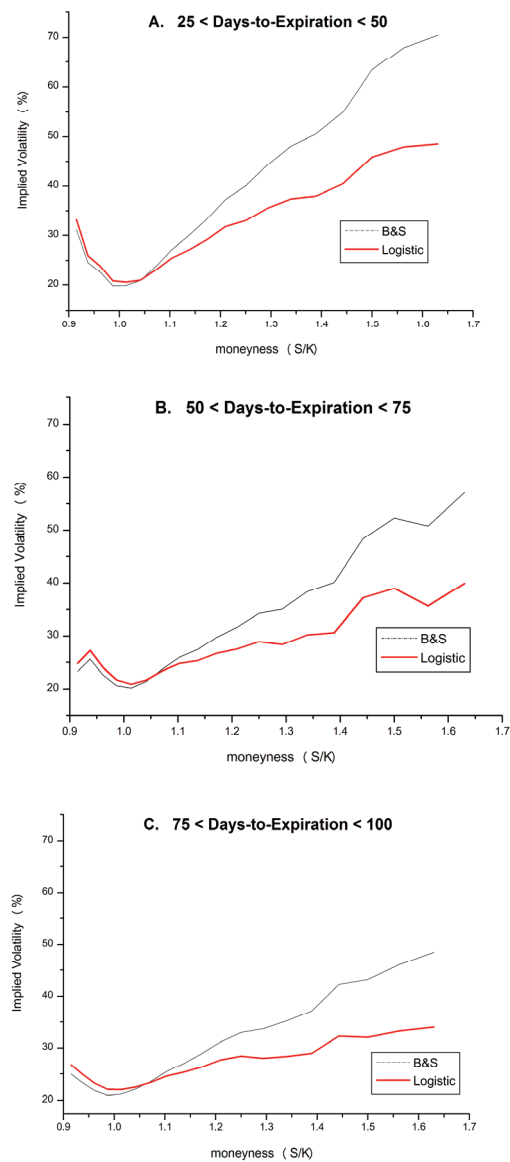


Figure 3. The implied volatility as a function of moneyness for B&S (light) and the logistic option pricing formula (bold). The volatility smile still exists with logistic option pricing, but it is much less pronounced than for B&S. Note that the moneyness range is much wider than ranges typically reported in the literature.

4.3. Hedging Errors

The logistic pricing formula implies a delta hedge given by Equation (11). We test the performance of this hedge ratio relative to the standard B&S delta hedging by shorting the call and hedging the position with Δ shares and the risk-free asset. We record the value of this zero-cost portfolio on the next day:

$$H = -C(t + 1) + \Delta(t)S(t + 1) + cash(t)e^{rdt},$$

(14)

where $cash(t) = C(t) - \Delta(t)S(t)$. According to the theoretical model with perfect hedging, H should be zero. Thus, H measures the empirical hedging error, i.e., the difference between the next-day value of the option and the replicating portfolio (see, for example, Bakshi et al. (1997)). As we are hedging many options with very different values, we look at the normalized error, which is the error H divided by the absolute size of the position (i.e., $C(t) - cash(t) = \Delta(t)S(t)$; note that $cash$ is negative). In other words, we measure the error for a position of USD 1. We analyze the error both for the B&S delta and for the logistic delta above. In both cases, we use the VIX as an estimate of the volatility.⁷ In the B&S formula, σ is the annual volatility; thus, we simply take $\sigma_{B\&S} = VIX$. In the logistic option pricing formulas, $\sigma_{Logistic}$ is the volatility of returns for the period until expiration (not annualized). We approximate this value by $\sigma_{Logistic} = VIX \cdot \sqrt{T}$, where T is the time to expiration in years. While the scaling by $\sqrt{T}$ is mathematically precise only for i.i.d. log-returns, it is a good approximation for returns (rather than log-returns) as well, at least for horizons exceeding one month—see Figure A1 in the Appendix A.

Table 4 reports the mean absolute normalized hedging error, as well as the mean square normalized hedging error for three different ranges of time to expiration. In all cases the logistic formula yields smaller hedging errors than B&S. The difference is significant at the 1% level in all cases. While the hedging errors, and hence the differences reported in the table, are small in absolute values, the relative improvement is rather large. The mean square normalized error of the logistic model is 15.5% lower than B&S for 25–50 days to expiration, 20% lower for 50–75 days to expiration, and 10.9% lower for 75–100 days to expiration. Bakshi et al. (1997) report that the hedging errors for the various stochastic volatility models are similar to those of B&S when a single instrument is used. Thus, it seems that the logistic model outperforms these models in single-instrument hedging. Extending the model and adding instruments may further reduce hedging errors.

The hedging error is given by $H = -C(t + 1) + \Delta(t)S(t + 1) + cash(t)e^{rdt}$. It is the difference between the next-day price of the option and the replicating portfolio. The normalized hedging error is given by the error, normalized by the absolute size of the position $H/\Delta(t)S(t)$, i.e., the error per USD 1 position. Both the absolute and the squared normalized hedging error are smaller for the logistic option pricing model than for B&S. This holds for all three time-to-expiration categories. In all cases the difference is significant at the 1% level.

Table 4. Hedging errors.

	25–50 Days to Expiration		50–75 Days to Expiration		75–100 Days to Expiration	
	B&S	Logistic	B&S	Logistic	B&S	Logistic
Mean absolute normalized hedging error ($\cdot 10^{-3}$)	2.79	2.64	2.94	2.76	2.47	2.41
Mean square normalized hedging error ($\cdot 10^{-5}$)	2.65	2.24	3.57	2.86	1.92	1.71

5. Possible Origin of the Logistic Return Distribution

Section 2 shows that the empirical return distribution is closer to the logistic distribution than to any of the other distributions examined for horizons of up to one year. This is the motivation for deriving an option pricing formula based on the logistic return distribution. For this purpose, which is the main contribution of the present paper, one does not necessarily need to know *why* the return distribution is approximately logistic. Still, it is an intriguing question. While we have no pretense of offering a definitive answer to this question, we would like to sketch a possible explanation.

The central limit theorem asserts that the sum of i.i.d. random variables with finite variance converges to the normal distribution. Thus, if we denote the one-period total return ($1 + \text{rate of return}$) with $\tilde{r}$, the T-period total return is given by $\tilde{R} = \tilde{r}_1 \cdot \tilde{r}_2 \cdots \tilde{r}_T$ and $\log(\tilde{R}) = \sum_{t=1}^T \log(\tilde{r}_t)$. If the $\tilde{r}$ values are i.i.d., then so are their logs; therefore, taking the central limit theorem as $T \rightarrow \infty$, $\log(\tilde{R})$ converges to the normal distribution, and thus $\tilde{R}$ converges to the log-normal distribution. In B&S's continuous time model, the $\tilde{r}$ values are instantaneous; therefore, the distribution of $\tilde{R}$ is log-normal for any finite time interval, no matter how short.

Empirically, this is clearly not the case, as shown in Section 2. One possible explanation is that the one-period returns are not independent. This may certainly be true. However, the central limit theorem also holds for many cases with serial correlation (Billingsley (2008); Bradley (2005); and Durrett (2010)). More importantly, it is not clear why serial correlation would lead specifically to the logistic distribution.

We would like to suggest an explanation which is not based on the violation of the i.i.d assumption, but rather of the continuous-time concept of an “instantaneous” return. It seems that there is some natural limit of a maximal frequency at which new information arrives, and that it is meaningless to talk about returns over shorter periods. The minimal time period can be a minute, a second, or even a millisecond, but it is not instantaneous.⁸ Therefore, even if returns are i.i.d. and with finite variance, it will take some time for the return distribution to converge to the log-normal. This may seem like a minor technical point, but in fact, we believe that it may be the basis for understanding the origin of the empirical logistic return distribution. This understanding is based on the following three facts:

- The empirical return distribution over very short periods (15 s to a minute) is very close to a truncated Lévy (stable Paretian) distribution with an exponent α of approximately 1.4 (Mantegna and Stanley 1995).
- This distribution is very similar to the logistic distribution. Figure 4 demonstrates this point by comparing these two distributions.
- It is well-known that the Lévy distribution is stable. When the Lévy distribution is truncated, it is pseudo-stable. This means that it eventually converges to the normal distribution because it has finite variance and the conditions of the central limit theorem hold; however, this convergence can take a *very long* time (Mantegna and Stanley 1994). For holding periods shorter than this convergence time, the distribution remains an approximately truncated Lévy distribution, and hence approximately logistic.

This is one possible way to think about the origin of the logistic return distribution. However, it only pushes the mystery one level deeper: why is the very short horizon return distribution a truncated Lévy distribution? Here the explanation becomes even more speculative. Levy (2005) suggests that the truncated Lévy distribution of returns is related to the power law tail distribution of wealth.⁹ He compares the empirical wealth distribution power law exponent with the Lévy return distribution exponent across different countries and finds close agreement between these values in each country (but a difference in these parameters across countries). Of course, other explanations for the short-horizon truncated Lévy return distribution are possible. We leave this line of inquiry for future work.

The truncated Lévy and logistic distributions are very similar. However, working with the truncated Lévy distribution is much more cumbersome, as there is no analytical expression for the p.d.f of this distribution. In contrast, the logistic distribution proves very easy to work with, as shown in Section 3.

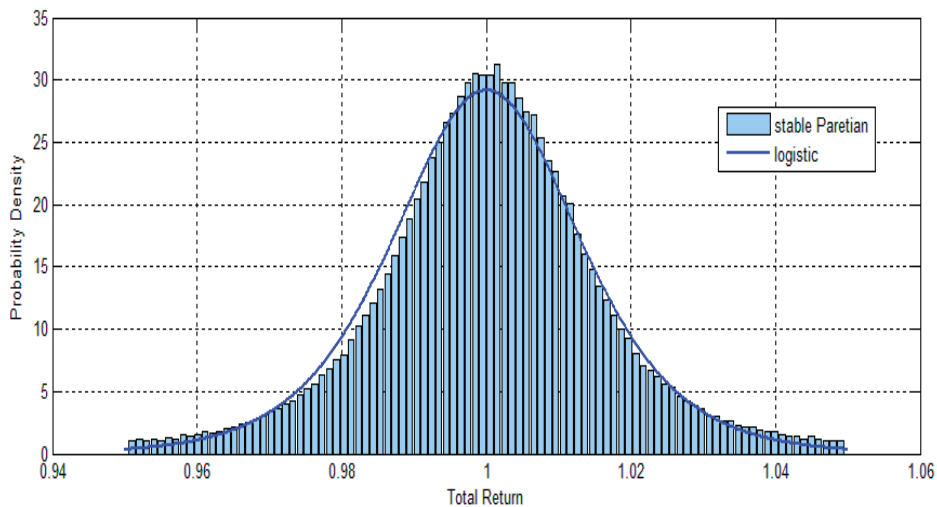


Figure 4. The logistic distribution and the truncated symmetric stable Paretian distributions are very close. The histogram is created by drawing 200,000 random observations from a truncated stable Paretian distribution with $\alpha = 1.4$, $\beta = 0$ (asymmetry parameter), $\gamma = 0.01$ (scale parameter), and $\delta = 1$ (location parameter). The solid line is the best logistic fit. The fit depends on the stable Paretian distribution being symmetric ($\beta = 0$), but it is not sensitive to the values of γ and δ . The excellent fit holds for a wide range of the exponent α .

6. Concluding Remarks

The standard approach to option pricing is to start with a continuous-time process for the evolution of the underlying asset's price and to derive option prices under this process. Great progress has been made with this approach. More and more sophisticated models incorporating time-varying volatility, jumps, and stochastic interest rates have been solved. These models typically improve empirical pricing and hedging errors relative to the cornerstone B&S model, but this comes at a price of introducing more parameters that are usually econometrically challenging to estimate. This approach typically leads to a conditional return distribution of the underlying asset that is either log-normal or approximately log-normal with small deviations due to infrequent jumps. However, for horizons of up to a year, the empirical return distribution is very different from the log-normal. Rather, it is much closer to the logistic distribution, which provides the best fit to the empirical distribution out of a wide range of theoretical distributions examined.

This observation is the motivation for the present paper, which takes a different approach to option pricing. Rather than starting with a continuous-time return process, our starting point is the assumption of a logistic distribution of the underlying asset value at expiration (i.e., a logistic return distribution). By employing risk-neutral valuation, we analytically derive option prices under the logistic distribution. From this valuation formula we also derive the corresponding hedge ratio.

The logistic option pricing formulas turn out to be very elegant and easy to implement. While the logistic distribution is similar to the stable Paretian distribution employed by Carr and Wu (2003), it is much easier to work with and yields a simple analytic pricing formula. The logistic option pricing formula involves exactly the same parameters as the B&S formula, with only one free parameter that needs to be estimated—the volatility. Empirically, the logistic option pricing formulas yield prices and hedge ratios which are much better than B&S and are roughly comparable to more complex models, involving a larger number parameters that are typically econometrically challenging to estimate.

We see this paper as a first step. Combining the logistic pricing approach with elements that have been documented to be important for option pricing, such as time-varying volatility, jumps, and possibly stochastic interest rates could be a promising next step. This would require explaining the logistic distribution as the result of an underlying short-term (but not continuous) return process with these elements. Clearly, there is much to do along these lines. Adding these elements should hopefully even further improve the pricing and hedging performance of the logistic option pricing model.

Author Contributions: Conceptualization, H.L.; Formal analysis, M.L.; Investigation, H.L. and M.L.; Writing—original draft, H.L. and M.L.; Writing—review & editing, M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data available from authors upon request.

Acknowledgments: We are grateful to the three anonymous referees for their many helpful comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

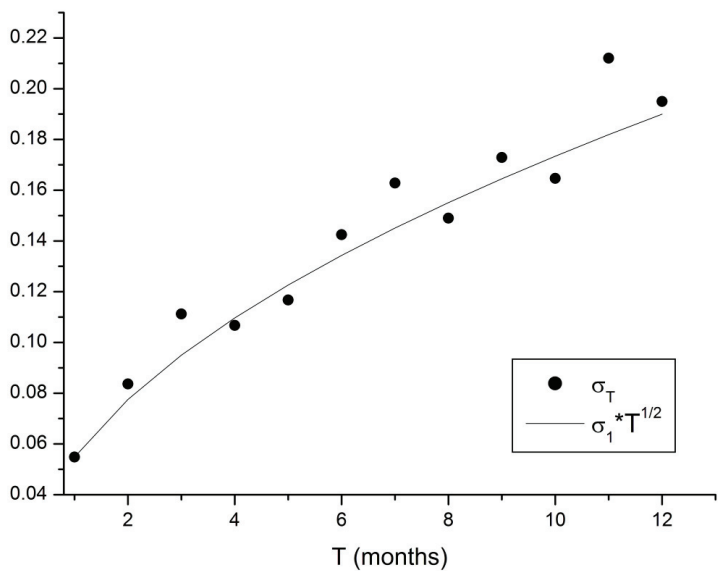


Figure A1. The sample is the set of S&P 500 monthly returns over 1926–2013. σ_1 is the empirical one-month standard deviation of returns. σ_T is the empirical standard deviation of returns for the T -month holding period. While the theoretical relation $\sigma_T = \sigma_1 \cdot \sqrt{T}$ holds mathematically only for i.i.d log-returns, empirically it provides a very good approximation for returns as well, at least for horizons exceeding one month.

- Notes**
- ¹ An important exception that does not imply a log-normal return distribution is the model of Carr and Wu (2003), who employ the stable Pareto return distribution to explain the observation that the volatility smirk does not flatten out as maturity increases. The logistic distribution employed in the present study is very similar to a truncated stable Paretian distribution (see Section 5 and Figure 4). The advantage of using the logistic distribution is that it yields a very simple analytic option pricing formula.
 - ² Raja (2009) reaches the same conclusion for individual U.K. stocks. He finds that the best fit for log-returns is provided by the log-logistic distribution, implying that returns are distributed logistically.

- ³ For the longer holding periods the conditional distribution analysis is both less relevant and less feasible. The typical length of volatility regimes is probably not very long. Practically, partitioning the 88 year sample into quintiles leads to less than 18 annual observations per quintile, which makes estimating the shape of the conditional distribution impractical. Finney (1978) shows that for a special choice of parameters, the logistic distribution is close to the normal distribution. This is employed by Hofstetter and Selby (2001) to obtain an analytic approximation for the B&S implied volatility. The results reported in this section show that the empirical return distribution is significantly different than the log-normal and normal distributions and is much closer to the logistic distribution. In other words, for the empirical distribution, the special parametrization under which the logistic and the normal are similar does not hold, and the two distributions are quite different.
- ⁴ If the asset's end-of-period value at maturity is distributed log-normally, the return distribution is also log-normal. This is true both for the physical return distribution and the risk-neutral return distribution, although these two distributions have different means. Similarly, a logistic end-of-period asset value implies a logistic risk-neutral return distribution.
- ⁵ The logistic distribution is a part of the family of elliptical distributions, of which the normal distribution is also a member. The CAPM holds for all distributions in this family (with the other standard assumptions; see Chamberlain (1983), Owen and Rabinovitch (1983), and Berk (1997)).
- ⁶ For practical purposes, it is convenient to have a simple way to translate annual volatility to the volatility for different horizons. While the scaling of the standard deviation with time by $\sqrt{T}$ is mathematically precise only for log-returns and under the assumption of no serial correlation, Figure A1 in the Appendix A shows that this scaling is a very good approximation for empirical returns (rather than log-returns) as well.
- ⁷ Similar results, although more noisy, are obtained when the option's implied volatility is used instead of the VIX.
- ⁸ For example, suppose that new information/liquidity shocks arrive every second, and this determines some 1 s return distribution. One could technically define a 1 ms return distribution for this process, but this would be meaningless—it would not change the 1 s return distribution dictated by Nature, nor induce it to be log-normal. For the advantages of the discrete-time approach see Brennan (1979).
- ⁹ The idea is based on the fact that the sum of i.i.d. random variables drawn from any distribution with a power law tail and exponent $\alpha < 2$ converges to the Lévy distribution with an exponent α . Thus, if an individual's influence on the stock price is proportional to her wealth, and the wealth distribution has a power law tail, then the return, which is the sum of the actions of many individuals, will be distributed according to the Lévy distribution with the same exponent as that of the wealth distribution. The truncation is due to the finiteness of the economy.

References

- Ackerer, Damien, and Damir Filipović. 2020. Option pricing with orthogonal polynomial expansions. *Mathematical Finance* 30: 47–84. [CrossRef]
- Amin, Kaushik I., and Robert. A. Jarrow. 1992. Pricing options on risky assets in a stochastic interest rate economy. *Mathematical Finance* 2: 217–37. [CrossRef]
- Amin, Kaushik I., and Victor. K. Ng. 1993. Option valuation with systematic stochastic volatility. *The Journal of Finance* 48: 881–910. [CrossRef]
- Bailey, Warren, and Rene. M. Stulz. 1989. The pricing of stock index options in a general equilibrium model. *Journal of Financial and Quantitative Analysis* 24: 1–12. [CrossRef]
- Bakshi, Gurdip, Charles Cao, and Zhiwu Chen. 1997. Empirical performance of alternative option pricing models. *The Journal of Finance* 52: 2003–49. [CrossRef]
- Barone-Adesi, Giovanni, Robert F. Engle, and Lorian Mancini. 2008. A GARCH option pricing model with filtered historical simulation. *Review of Financial Studies* 21: 1223–58. [CrossRef]
- Bartl, Daniel, Samuel Drapeau, and Ludovic Tangpi. 2020. Computational aspects of robust optimized certainty equivalents and option pricing. *Mathematical Finance* 30: 287–309. [CrossRef]
- Bates, David S. 1991. The crash of '87: Was it expected? The evidence from options markets. *The Journal of Finance* 46: 1009–44. [CrossRef]
- Becker, Garry F. 1931. *Hyperbolic Functions*. Redditch: Read Books.
- Berk, Jonathan B. 1997. Necessary conditions for the CAPM. *Journal of Economic Theory* 73: 245–57. [CrossRef]
- Billingsley, Patrick. 2008. *Probability and Measure*. Hoboken: John Wiley & Sons.
- Black, Fischer, and Myron Scholes. 1973. The pricing of options and corporate liabilities. *The Journal of Political Economy* 81: 637–54. [CrossRef]
- Bradley, Richard C. 2005. Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys* 2: 107–44. [CrossRef]
- Brennan, Michael. 1979. The pricing of contingent claims in discrete time models. *The Journal of Finance* 24: 53–68.
- Carr, Peter, and Liuren Wu. 2003. The finite moment log stable process and option pricing. *The Journal of Finance* 58: 753–78. [CrossRef]
- Chamberlain, Gary. 1983. A characterization of the distributions that imply mean—Variance utility functions. *Journal of Economic Theory* 29: 185–201. [CrossRef]

- Cox, John C., Stephen A. Ross, and Mark Rubinstein. 1979. Option pricing: A simplified approach. *Journal of Financial Economics* 7: 229–63. [CrossRef]
- Cremers, K. Marijn, Joost Driessen, and Pascal Maenhout. 2008. Explaining the level of credit spreads: Option-implied jump risk premia in a firm value model. *Review of Financial Studies* 21: 2209–42. [CrossRef]
- Duffie, Darrell, Jun Pan, and Kenneth Singleton. 2000. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68: 1343–76. [CrossRef]
- Durrett, Rick. 2010. *Probability: Theory and Examples*. Cambridge: Cambridge University Press.
- Eraker, Bjorn. 2004. Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *The Journal of Finance* 59: 1367–404. [CrossRef]
- Fama, Eugene. F. 1965. The behavior of stock-market prices. *Journal of Business* 38: 34–105. [CrossRef]
- Finney, David J. 1978. *Statistical Methods in Biological Assay*, 3rd ed. London: Griffin.
- Geske, Robert. 1979. The valuation of compound options. *Journal of Financial Economics* 7: 63–81. [CrossRef]
- Heston, Steven L. 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* 6: 327–43. [CrossRef]
- Hofstetter, Etienne, and Michael J. Selby. 2001. *The Logistic Function and Implied Volatility: Quadratic Approximation and Beyond*. Working Paper. Coventry: Warwick Business School.
- Hull, John. 2009. *Options, Futures and Other Derivatives*. London: Pearson Education.
- Hull, John, and Alan White. 1987. The pricing of options on assets with stochastic volatilities. *The Journal of Finance* 42: 281–300. [CrossRef]
- Kirkby, J. Lars, and Duy Nguyen. 2020. Efficient Asian option pricing under regime switching jump diffusions and stochastic volatility models. *Annals of Finance* 16: 307–51. [CrossRef]
- Kou, Steven G. 2002. A jump-diffusion model for option pricing. *Management Science* 48: 1086–101. [CrossRef]
- Levy, Haim, and Ran Duchin. 2004. Asset return distributions and the investment horizon. *The Journal of Portfolio Management* 30: 47–62. [CrossRef]
- Levy, Moshe. 2005. Market Efficiency, the Pareto Wealth Distribution, and the Lévy Distribution of Stock Returns. In *The Economy as an Evolving Complex System, III: Current Perspectives and Future Directions*. Oxford: Oxford Academic, p. 101.
- Madan, Dilip B., Peter P. Carr, and Eric C. Chang. 1998. The variance gamma process and option pricing. *European Finance Review* 2: 79–105. [CrossRef]
- Mandelbrot, Benoit B. 1963. The Variation of Certain Speculative Prices. *Journal of Business* 36: 394–419. [CrossRef]
- Mantegna, Rosario N., and Eugene H. Stanley. 1994. Stochastic process with ultraslow convergence to a Gaussian: The truncated Lévy flight. *Physical Review Letters* 73: 2946. [CrossRef] [PubMed]
- Mantegna, Rosario N., and Eugene H. Stanley. 1995. Scaling behaviour in the dynamics of an economic index. *Nature* 376: 46–49. [CrossRef]
- Melino, Angelo, and Stuart M. Turnbull. 1990. Pricing foreign currency options with stochastic volatility. *Journal of Econometrics* 45: 239–65. [CrossRef]
- Merton, Robert C. 1973. Theory of rational option pricing. *The Bell Journal of Economics and Management Science* 4: 141–83. [CrossRef]
- Merton, Robert C. 1976. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics* 3: 125–44. [CrossRef]
- Naik, Vasanttilak. 1993. Option valuation and hedging strategies with jumps in the volatility of asset returns. *The Journal of Finance* 48: 1969–84. [CrossRef]
- Officer, Robert R. 1972. The distribution of stock returns. *Journal of the American Statistical Association* 67: 807–12. [CrossRef]
- Owen, Joel, and Ramon Rabinovitch. 1983. On the class of elliptical distributions and their applications to the theory of portfolio choice. *The Journal of Finance* 38: 745–52. [CrossRef]
- Pan, Jun. 2002. The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics* 63: 3–50. [CrossRef]
- Raftery, Adrian E. 1995. Bayesian model selection in social research. *Sociological Methodology* 25: 111–63. [CrossRef]
- Raja, Ammar. 2009. *Future of Option Pricing: Use of Log Logistic Distribution Instead of Log Normal Distribution in Black Scholes Model*. MPRA Working Paper. Munich: University Library of Munich.
- Scott, Louis O. 1987. Option pricing when the variance changes randomly: Theory, estimation, and an application. *Journal of Financial and Quantitative Analysis* 22: 419–38. [CrossRef]
- Stein, Elias M., and Jeremy C. Stein. 1991. Stock price distributions with stochastic volatility: An analytic approach. *Review of Financial Studies* 4: 727–52. [CrossRef]
- Wiggins, James B. 1987. Option values under stochastic volatility: Theory and empirical estimates. *Journal of Financial Economics* 19: 351–72. [CrossRef]
- Willems, Sander. 2019. Asian option pricing with orthogonal polynomials. *Quantitative Finance* 19: 605–18. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

Can Investment Views Explain Why People Insure Their Cell Phones But Not Their Homes?—A New Perspective on the Catastrophe Insurance Puzzle

Annette Hofmann ^{1,*} and Peter Zweifel ²

¹ Department of Finance, Lindner College of Business, University of Cincinnati, Cincinnati, OH 45221, USA

² Department of Economics, Emeritus, University of Zurich, CH-8001 Zürich, Switzerland; peter.zweifel@uzh.ch

* Correspondence: annette.hofmann@uc.edu

Abstract: The consistently missing demand for catastrophe insurance and for coverage of other low-probability–high-consequence risks is often referred to as the catastrophe insurance puzzle. People show reluctance to insure low-probability–high-consequence events, even with some disastrous consequences, yet insure against small high-probability–low-consequence events. There has been no convincing explanation of this puzzle to this date. This article points out that the underlying rationale may be that individuals interpret insurance contracts with low payout probability as an investment with negative expected net present value. While premium payments start with the conclusion of the contract, usually there is only one loss payment in the near or far future. Using a simple annuity model with fixed annual premiums and expected indemnity payouts, it is found that even an individual characterized by the degree of risk aversion found in the literature is unlikely to purchase insurance with these characteristics. To alleviate this unfavorable insurance purchase syndrome, combining a low-probability with a high-probability loss insurance contract may be a way to incentivize individuals to purchase catastrophe risk coverage.

Keywords: catastrophe insurance puzzle; insurance demand for low-probability events; investment view; risk aversion; biases in risk perception

Citation: Hofmann, Annette, and Peter Zweifel. 2024. Can Investment Views Explain Why People Insure Their Cell Phones But Not Their Homes?—A New Perspective on the Catastrophe Insurance Puzzle.

Journal of Risk and Financial Management 17: 30. <https://doi.org/10.3390/jrfm17010030>

Academic Editor: Thanasis Stengos

Received: 21 December 2023

Revised: 5 January 2024

Accepted: 7 January 2024

Published: 12 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Studies of insurance demand often suggest that individuals seem to either ignore or at least undervalue low-probability events. For rare but high-impact events, insurance is often not attractive. The consistently missing demand for catastrophe insurance and coverage of other low-probability–high-consequence risks is often referred to as the *catastrophe insurance puzzle* (CIP). People show reluctance to insure low-probability–high-consequence events even if they have disastrous consequences, yet insure against small high-probability losses.

A possible explanation for this observation may be that individuals tend to pay little attention to risks with a probability of occurrence below some perceived threshold (Kunreuther and Pauly 2004; Slovic et al. 1977). This threshold may serve as a gatekeeper, guiding individuals to purchase insurance only for risks that are worthwhile to insure (those perceived to happen in the near future). As first mentioned by Kunreuther and Slovic (1978), an explanation of this behavior may also be that insurance is interpreted as some form of an investment. Indeed, insuring against rare hazards may be a bad investment most of the time. As Kunreuther and Slovic (1978, p. 66) state,

“there is evidence that people do not voluntarily insure themselves against natural disasters even when the rates are highly subsidized. The reasons for failure of insurance markets need to be understood, as they have important implications for policy.”

Interestingly, this early explanation from 1978 has never been studied in detail or even been evaluated more carefully in the literature. It is the research question of this article. What are the consequences of individuals interpreting insurance contracts with low payout probability as an investment with negative expected net present value but those with high payout probability as an investment with positive expected net present value? Using a simple annuity model with fixed annual premiums and expected indemnity payouts, it turns out that even an individual characterized by a degree of risk aversion found in the literature is unlikely to purchase insurance covering a low-probability-high-consequence event. However, combining a low-probability with a high-probability loss insurance contract may be a way to incentivize individuals to purchase catastrophe risk coverage, thereby alleviating the catastrophe insurance puzzle.

This article is structured as follows. Section 2 discusses the literature on the CIP and related questions. The subsequent section deals with the case of corporate insurance, where risk aversion does not play a role if management acts in the interest of well-diversified owners of the firm (Doherty 1985; Zweifel et al. 2021, chapter 5). Section 4 is devoted to the case of a consumer characterized by risk aversion and willing to pay a risk premium for a high-probability-low-consequence policy (exemplified by cell phone insurance). Section 5 pursues the idea of combining the former with a low-probability-high-consequence policy to overcome the behavior giving rise to the CIP. The last section discusses findings and draws conclusions.

2. Related Literature

Several possible explanations of the CIP exist in the literature but none seems convincing enough to have established itself as the main CIP rationale so far. First, it has been argued that the high transaction costs in catastrophe insurance markets may hinder demand (Kunreuther and Pauly 2004; Akter et al. 2008; Pothon et al. 2019). Kunreuther and Pauly (2004) studied a modeling approach to the purchase of catastrophe insurance and argued that a reason for the non-purchase of such insurance coverage may be the insufficient magnitude of the potential loss relative to its probability and to the insurance premium. Since the probability is so low, the severity would need to be more than bankruptingly high to warrant purchase.

Slovic et al. (1977) note that, in order to change individuals' perspectives and possibly motivate them to purchase insurance for rare losses (Slovic et al. 1977, p. 256),

“instead of describing the chances of a 100-year flood as 0.01 per year, one could note that an individual living in a particular house for 25 years faces a 0.22 chance of suffering 100-year damage at least once.”

Akter et al. (2008) study the uptake of catastrophe insurance among 3000 residents in six different districts in Bangladesh facing various levels of exposure. They identify not only common misunderstandings of risk but also discuss the high cost of insurance for the poor as factors limiting the demand for coverage. Among the reasons for going without coverage, ‘limited financial income’ and ‘dislike of terms and conditions’ figured prominently.¹ Pothon et al. (2019) evaluate twenty years of data in their attempt at understanding why people do not purchase earthquake insurance. They find that homeowners do understand the risk but consider the annual premium ‘too high’: on average, they would have to decrease from USD 980 to USD 160 annually to be acceptable.

Second, information asymmetries such as adverse selection and moral hazard may lead to market failure (Antwi-Boasiako 2014; Lin 2020). Antwi-Boasiako (2014) discusses market failures including adverse selection, moral hazard, correlated risks, and time inconsistency as potential reasons for why people do not purchase catastrophe insurance. Lin (2020) demonstrates that there is a positive demand-risk correlation in the earthquake insurance market in California, a finding consistent with adverse selection.

Third, limited liability and anticipation of government relief may inhibit catastrophe insurance demand. Governments often feel a need for the provision of relief in the case of a major catastrophe ('Samaritan's Dilemma'). Andor et al. (2020) evaluate the issue that, in the aftermath of disasters, governments often offer financial assistance for affected households. If anticipated by households, this introduces a crowding-out effect on insurance demand and possibly even precautionary ex ante measures. This crowding-out effect is sometimes referred to as "charity hazard". While empirical studies offer mixed results, Andor and colleagues find a positive correlation between governmental relief and non-financial protection measures.

Last, behavioral biases may contribute to why people do not purchase low-probability event insurance (Kunreuther and Pauly 2004; Antwi-Boasiako 2014).² Kunreuther and Pauly (2004) argue that a possible reason is that consumers may not know or may be unable to evaluate and understand low loss probabilities. For instance, when psychological biases are at play, under- or overweighting of small probabilities may hinder demand for coverage. Kunreuther and Pauly (2004) point out that people tend to buy (more) catastrophe insurance right after a major disaster, when they revise their subjective probability estimates ('it does happen!'). This behavior is known as *salience bias*, meaning a focus on items or information that grab people's attention. In addition, the high deductibles common in catastrophe insurance make consumers feel a lack of financial protection. Indeed, rich context information must be made available to individuals to enable them to judge differences between low probabilities. The authors find that individuals need to be presented with "comparison scenarios that are located on the probability scale to evoke people's own feelings of risk". Individuals may indeed have difficulty processing very low probabilities or even fail to distinguish between low-probability and zero-probability events, and the missing demand for insurance covering low-probability events may indeed be due to this ignoring of risks with low probabilities.³ As shown in several experiments and field surveys, individuals consistently buy insurance only when the probability of loss is above a certain threshold (Slovic et al. 1977; Kunreuther and Slovic 1978), even in cases where the premium is heavily subsidized. To our knowledge, there is only one experiment where this relationship is reversed (Laury et al. 2009).

However, all of these possible explanations in the literature of the CIP fail to convincingly explain the consistent lack of demand for insurance covering low-probability-high-consequence events *and* high demand for insurance covering high-probability-low-consequence events. It is this combination of phenomena that constitutes the *catastrophe insurance puzzle* (CIP). The objective of this contribution is to explain the CIP using a simple theoretical model that takes into account the time value of money while maintaining the assumption of a rational risk-neutral or risk-averse decision maker.

A case in point for this combined puzzle is homeowners' insurance against flooding, which is subsidized up to 50 percent (Brown 2011). Yet, after each major disaster, the question of why homeowners do not adequately insure against catastrophic losses comes up in the media again. Interestingly, a recent study by Browne et al. (2015) uses data from an insurer that offers add-on coverage for both a low-probability-high-consequence risk (a flood peril) and a high-probability-low-consequence risk (a bicycle theft). They find evidence consistent with the combined CIP in the sense that a significantly higher number of policyholders purchase add-on coverage to their homeowner's insurance to cover the risk of bicycle theft than to cover the risk of major loss due to flooding.

Another particularly striking instance of the combined CIP is the growing market for cell phone insurance in the presence of the continuing under-insurance of flooding risks. According to a recent report, the global mobile phone insurance market was estimated at USD 24.6 bn in 2020, with a growth rate of 11.8% for the following decade.⁴ It is expected to reach USD 38.1 bn by 2024. This growth is puzzling because damage to or loss of a cell phone entails getting accustomed to a new phone in the extreme case, an event with minimal impact on life, contrary to most low-probability-high-impact losses such as natural and other disasters.

In the following, an explanation of this and other instances of the combined CIP is proposed which does not have to invoke any of the behavioral deviations from rationality cited above. It simply views insurance as a financial investment with expected returns that may be frequently negative.

3. The Corporate View of Insurance as an Investment

3.1. Low-Probability Insurance

The purchase of insurance is viewed as an investment with a *risky* net present value (PV) of returns, assuming no risk aversion, thus adopting the point of view of a risk neutral individual or a corporation. As a first step, a general result is derived (a realistic value for n , the number of years, will be introduced below). The per-period premium is P_I for coverage I of a low-probability–high-consequence type of event with loss size L . With a real rate of discount $r < 1$, the (certain) PV of the premium stream is

$$PV_I(Premia) = P_I + \frac{P_I}{(1+r)} + \frac{P_I}{(1+r)^2} + \dots + \frac{P_I}{(1+r)^n} \quad (1)$$

Multiplying Equation (1) by $(1+r)$ and subtracting the result from Equation (1), one obtains

$$(1+r)PV_I(Premia) - PV_I(Premia) = r \cdot PV_I(Premia) = P_I - \frac{P_I}{(1+r)^n},$$

resulting in a standard annuity present value,

$$PV_I(Premia) = \frac{P_I}{r} - \frac{P_I}{r(1+r)^n} \quad (2)$$

Next, the expected PV of the benefit stream associated with a low-probability–high-consequence event is calculated. Given a binary distribution with (low) constant probability ρ_I , the expected waiting period Em for the first (and by assumption only) loss to materialize is given by $Em := 1/\rho_I$; before that, there are no benefits.⁵ Assuming full coverage, one obtains the expected PV

$$EPV_I(Benefits) = 0 + 0 + \dots + \frac{\rho_I L_I}{(1+r)^m} + 0 + 0 \dots = \frac{\rho_I L_I}{(1+r)^{1/\rho_I}} \quad (3)$$

The high value of L_I is weighted by a small value of ρ_I and strongly discounted because the expected waiting period $1/\rho_I$ is long in principle. Neglecting risk aversion and assuming that the premium is fair and the loss is fully insured ($I = L$), the net EPV of buying catastrophic insurance amounts to

$$\begin{aligned} EPV_I(Benefits) - PV_I(Premia) &= \frac{\rho_I L_I}{(1+r)^m} - \left(\frac{P_I}{r} - \frac{P_I}{r(1+r)^n} \right) \\ &= \frac{\rho_I L_I}{(1+r)^m} - \left(\frac{\rho_I L_I}{r} - \frac{\rho_I L_I}{r(1+r)^n} \right) \\ &= \rho_I L_I \left(\frac{1}{(1+r)^m} - \frac{1}{r} + \frac{1}{r(1+r)^n} \right) \\ &= \rho_I L_I \left(\frac{r(1+r)^n - (1+r)^{m+n} + (1+r)^m}{r(1+r)^{m+n}} \right) < 0 \end{aligned} \quad (4)$$

due to $(1+r)^{m+n} > r(1+r)^n + (1+r)^m$. Therefore, in a realistic scenario, catastrophe insurance constitutes a bad investment with a negative expected present value (NPV). Note that this holds for all arbitrary (positive) values of ρ_I , L_I , m , n , and $r < 1$ and for all types of premium because the actuarially fair premium is the lowest possible, without any loading. Given the loss size has no impact here, the findings hold for both low-severity and high-severity loss insurance.

3.2. High-Probability Insurance

For maximum contrast, now assume that a loss J occurs (and a benefit is paid) in every period with probability $\rho_J > \rho_I$ but with loss amounting to $L_J < L_I$, characterizing the “high-probability-small-consequence characteristic” of the policy. With the average duration of a policy equal to four years, the EPV_J of the benefit stream is given by⁶

$$EPV_J(\text{Benefits}) = \rho_J L_J + \frac{\rho_J L_J}{(1+r)} + \frac{\rho_J L_J}{(1+r)^2} + \frac{\rho_J L_J}{(1+r)^3} \quad (5)$$

Multiplying by $(1+r)$, one obtains

$$(1+r)EPV_J(\text{Benefits}) = (1+r)\rho_J L_J + \rho_J L_J + \frac{\rho_J L_J}{(1+r)} + \frac{\rho_J L_J}{(1+r)^2}$$

Subtracting this from Equation (5) and dividing by r yields

$$EPV_J(\text{Benefits}) = \rho_J L_J \left(\frac{1+r}{r} - \frac{1}{r(1+r)^3} \right) \quad (6)$$

Since the premium stream starts in period no. 1 when the first loss occurs, it has the same PV as the EPV of the benefit stream, resulting in

$$EPV_J(\text{Benefits}) - PV_J(\text{Costs}) = 0. \quad (7)$$

Therefore, without the effect of risk aversion, investment in a stand-alone high-probability-low-consequence insurance just breaks even regardless of loss size. This result is in accordance with the basic model of insurance demand, which predicts zero willingness to pay (WTP) in the absence of risk aversion. Indeed, for risk-neutral parties, the net present value of an insurance contract is zero.

3.3. Combining Low-Probability and High-Probability Insurance

For a policy combining coverage of a low-probability loss I and a high-probability loss J , a natural simplifying assumption is uncorrelatedness of the two risks. This results in an addition of payments and premiums, yielding in a negative value of the investment as before because zero is added to Equation (4) according to Equation (7). Again, this result holds for arbitrary values of ρ_I , L_I , m , n , and $r < 1$ and all types of premium.

Conclusion 1: From a corporation’s perspective, being risk neutral, a low-probability loss insurance contract is unambiguously considered a bad investment and will not be undertaken regardless of loss size. A stand-alone high-probability insurance contract just breaks even, again regardless of loss size. Combining low- and high-probability insurance still results in a bad investment.

4. The Individual View of Insurance as an Investment

4.1. Low-Probability Insurance

While for a risk-neutral party, WTP to get rid of a risk (in excess of expected value) is zero, it is positive for risk-averse parties. It amounts to the risk premium, defined as the difference between expected wealth and the (lower) certainty equivalent of wealth (Schlesinger and Venezian 1986; Zweifel et al. 2021, chapter 2). Evidently, the risk premium increases with the individual’s degree of risk aversion; however, the precise relationship depends on the higher-order moments of the wealth distribution (Stapleton and Zeng 2018). The expected net present value of the benefits now includes the risk premium denoted by $\pi_k > 0$, reflecting the individual’s benefit of getting rid of the risk completely by paying the fair competitive premium P_I for k periods. As before, the average duration of a policy

is equal to four years. Since the benefit of certainty has to be discounted to PV as well, Equation (3) becomes

$$\begin{aligned} EPV_I(Benefits) &= \pi_k + \frac{\pi_k}{1+r} + \frac{\pi_k}{(1+r)^2} + \frac{\pi_k}{(1+r)^3} + \frac{\rho_I L_I}{(1+r)^{1/\rho_I}} \\ &= \pi_k \left(\frac{1}{r} - \frac{1}{r(1+r)^3} \right) + \frac{\rho_I L_I}{(1+r)^{1/\rho_I}}. \end{aligned} \quad (8)$$

As to the premium stream, Equation (2) does not need to be modified because the premia can arguably be assumed to be risk-free. Combining Equations (8) and (9), one obtains

$$EPV_I(Benefits) - PV_I(Premia) = \pi_i \left(\frac{1}{r} - \frac{1}{r(1+r)^3} \right) + \frac{\rho_I L_I}{(1+r)^{1/\rho_I}} - P_I \left(\frac{1}{r} - \frac{1}{r(1+r)^3} \right), \quad (9)$$

which can be rewritten to become

$$\begin{aligned} EPV_I(Benefits) - PV_I(Premia) &= \frac{\rho_I L_I}{(1+r)^{1/\rho_I}} - (P_I - \pi_k) \left(\frac{1}{r} - \frac{1}{r(1+r)^3} \right) \\ &= \frac{\rho_I L_I}{(1+r)^{1/\rho_I}} - P_I \left(\frac{1}{r} - \frac{1}{r(1+r)^3} \right) + \pi_k \left(\frac{1}{r} - \frac{1}{r(1+r)^3} \right). \end{aligned} \quad (10)$$

From Equation (4), one has

$$\frac{\rho_I L_I}{(1+r)^{1/\rho_I}} - P_I \left(\frac{1}{r} - \frac{1}{r(1+r)^3} \right) < 0,$$

implying that the sum of the first two terms of Equation (10) is negative. Therefore, catastrophe insurance constitutes a good investment with a positive expected PV only as long as the individual is sufficiently risk-averse to exhibit a high-risk premium π_k .

Equation (10) can be evaluated further by inserting, e.g., $r = 0.05$ and $1/\rho_I = 20$. This means that, in Equation (10), the first term is multiplied by $1/2.65 = 0.377$ and the second and third by $20 - 17.38 = 2.72$. Given fair premiums, $P_I = \rho_I L_I$, so the sum of the negative terms amounts to $2.343P_I$, but the positive last term amounts to $2.72\pi_k$. Therefore, the risk premium would have to be at least as large as 86% of the actuarially fair premium for low-probability-high-consequence insurance to constitute a good investment. However, in their review of estimates, Fracasso et al. (2023) report many instances where the risk premium is not significantly different from zero, making satisfaction of this condition unlikely. Yet it is known that the subjective rate of discount exceeds the market rate by far. A field experiment involving the general population in Denmark yielded a value of 28% Harrison et al. (2002). Using $r = 0.20$, $1/\rho_I = 20$, and $n = 4$ results in multipliers of $1/3.83 = 0.261$ and $5 - 17.38 = 2.72$, respectively.

Conclusion 2: From the perspective of risk-averse individuals, whether a low-probability-high-consequence insurance contract typical of catastrophe insurance is considered a good or bad investment importantly depends on their subjective rate of discount. For those with a rate close to the market rate, it constitutes a bad investment, while for those with a rate close to that found in the general population, it may be a good one.

The economic reason is that a high rate of subjective discount moves a future loss payment forward to the present.

4.2. High-Probability Insurance

The average time before damage to or loss of a cell phone occurs is a mere 10 weeks in the United States, making it an archetypical high-probability-low-consequence event. Globally, the market of cell phone insurance was estimated at USD 24.6 bn in 2020, with a growth rate of 11.8% for the following decade.⁷ This dwarfs the market for catastrophe insurance with its USD 5.4 mn as of 2022 and a predicted 7% rate of growth until 2030.⁸

To explain this striking contrast, assume again that a loss J occurs (and a benefit is paid) in every period with probability $\rho_J > \rho_I$ but with loss amounting to $L_J < L_I$, characterizing the “high-probability-small-consequence characteristic” of the policy. In contrast to Section 3.1, $Em = 1/5$ reflects the finding cited above, implying $1/\rho_J = 5$ on an annual basis. However, most insurers cap coverage at two payments per year, while the average duration of a contract is about $n = 1.5$ years (<https://www.nerdwallet.com/article/finance/surprising-things-cell-phone-insurance>, accessed 2 December 2023). Finally, they impose a deductible of up to USD 299. Estimating its average value at USD 150, one obtains

$$EPV_J(\text{Benefits}) = \left(1 + \frac{1}{1+r/2}\right) 2(L_J - 150 + \pi_k) \quad (11)$$

Maintaining the assumption of fair premiums for consistency (which are paid annually in advance), their PV is given by

$$PV_J(\text{Premia}) = 2(L_J - 150) \left(1 + \frac{1}{1+r}\right) \quad (12)$$

In all, Equations (4) and (5) result in

$$\begin{aligned} EPV_J(\text{Benefits}) - PV_J(\text{Premia}) &= \left(1 + \frac{1}{1+r/2}\right) 2(L_J - 150 + \pi_k) - 2(L_J - 150) \left(1 + \frac{1}{1+r}\right) \\ &= \left(\frac{1}{1+r/2} - \frac{1}{1+r}\right) 2(L_J - 150) \\ &\quad + 2\pi_k(L_J - 150) \left(1 + \frac{1}{1+r/2}\right) > 0 \end{aligned} \quad (13)$$

Conclusion 3: A high-probability-low-consequence insurance contract, typical of damage or loss of a cell phone, is a good investment regardless of the individual’s risk premium and hence the degree of risk aversion, provided the deductible is fully accounted for, as in the context of a fair premium. The reason is that the loss event occurs probably within weeks after the purchase of the policy on average.

This result gives rise to the question of why insurers offer this kind of policy. One reason may be that they charge a high loading, contrary to the assumption of a fair policy; another, that they see the opportunity to be in frequent contact with customers, with the associated possibility of selling coverage of other risks.

5. Combining Low-Probability and High-Probability Insurance

The idea of combining insurance policies to boost WTP for catastrophe insurance is not entirely new. When the launch of all-hazard policies covering catastrophic events is discussed, Conclusion 3 points to the importance of having a high-probability component in such a combined policy. Moreover, the size of the risk premium is crucial for its low-probability component, as stated in Conclusion 2. A first approach is to simply scale up the two risk premia to obtain the combined WTP value, in spite of the finding of Stapleton and Zeng (2018) that all moments of the wealth distribution enter the determination of the risk premium. In Section 4.2, this problem is avoided by applying the Arrow-Pratt (1964) formula for approximately deriving the maximum WTP for certainty. In return, the fact that, for individual consumers, the values of ρ_I , ρ_J , L_I and L_J are estimates and hence subject to uncertainty can be accounted for.

5.1. Combining the Risk Premia for Low-Probability and High-Probability Insurance

In this section, $\pi_{I,k}$ and $\pi_{J,k}$ denote the risk premium for low-probability-high-loss insurance and for high-probability-low-consequence insurance, respectively. From Equations (3) and (6), one has

$$\begin{aligned}
 EPV_I(\text{Benefits}) - PV_I(\text{Premia}) + EPV_J(\text{Benefits}) - PV_J(\text{Premia}) \\
 = \frac{\rho_I L}{(1+r)^{1/\rho_I}} - P_I \left(\frac{1}{r} - \frac{1}{r(1+r)^3} \right) + \pi_{k,I} \left(\frac{1}{r} - \frac{1}{r(1+r)^3} \right) \\
 + \left(\frac{1}{1+r/2} - \frac{1}{1+r} \right) 2(L_J - 150) + 2\pi_{k,J} (L_J - 150) \left(1 + \frac{1}{1+r/2} \right) > 0
 \end{aligned} \quad (14)$$

almost certainly. The sum of the first three terms was found to be positive for $r = 0.20$, $1/\rho_I = 20$, and $n = 4$ provided $\pi_{k,I}$ is at least 10% of the premium P_I in Section 3.1, while the two last terms are positive for all values of r and $\pi_{k,J}$.

5.2. WTP for Certainty for Low-Probability and High-Probability Insurance Combined

In view of the findings of Stapleton and Zeng (2018), the maximum WTP value is derived using the Arrow-Pratt (1964) approximation, which does not depend on the higher-order moments of the wealth distribution. The WTP for certainty is given by $(1/2)Var(W)R_A$, where $Var(W)$ denotes the variance of the wealth distribution (which amounts to the variance of the loss if wealth is not subject to a background risk) and R_A , the coefficient of absolute risk aversion. However, estimates of R_A are few and far between. An early contribution by Friedman (1974) obtains $R_A = 3 \cdot 10^{-3}$ in the context of choices of health insurance policies (another high-probability type) in the United States. Using a negative exponential utility function (which makes R_A independent of wealth), Aditto (2011) comes up with values between $1.44 \cdot 10^{-5}$ and $1.35 \cdot 10^{-4}$ among farmers in Thailand. This means that estimates of R_A may diverge by as much as a factor of $2.1 \cdot 10^2 = 210$.

To avoid this problem, one may resort to the coefficient of relative risk aversion $R_R := W \cdot R_A$, with W denoting wealth, whose estimates differ much less. This is the appropriate risk measure (abstracting from higher-order measures such as prudence and temperance) if losses vary in proportion with wealth, which is the case for both the low-probability perils typical of catastrophe insurance and the high-probability ones typical of cell phone insurance, which both increase with income and wealth (Moscone and Tosetti 2010). From R_R , $R_A = R_R/W$ can be determined.

Meyer and Meyer (2005) note that a good deal of the divergence in estimates of R_R stems from different definitions of wealth (consumption, respectively) at risk. Using the estimates by Friend and Blume (1975) as their benchmark, they distinguish three types of wealth: (1) assets that are easily reallocated, (2) assets including housing, and (3) assets including housing as well as human capital. In the present context, category (2) is appropriate for the low-probability risk I and category (3) for the high-probability risk J (which is defined as healthcare expenditure here). After a series of adjustments, Meyer and Meyer (2005), Table 1 arrive at $R_R = 2.25$ for category (2) and $R_R = 2.82$ for category (3) in the USD 200,000-to-500,000 wealth bracket which is relevant in this context (see below). For simplicity, a common value $R_R = 2.5$ is used. As to wealth W , the mean value of real estate property in Florida's coastal communities was USD 340,000 as of 2021, while median non-housing wealth was USD 416,000 in 2020 (Hays and Sullivan 2022).⁹

For calculating $Var(\rho_I L_I) + Var(\rho_J L_J)$ (with the two risks uncorrelated and the deductibles imposed on both types of insurance for simplicity), one needs to have the variance of a product because consumers know ρ_I , ρ_J , L_I and L_J only approximately. This variance is given by

$$\begin{aligned}
 Var(\rho_I L_I) + Var(\rho_J L_J) &= \{Var(\rho_I) - E^2(\rho_I)\} \{Var(L_I) - E^2(L_I)\} - \{E(\rho_I)E(L_I)\}^2 \\
 &+ \{Var(\rho_J) - E^2(\rho_J)\} \{Var(L_J) - E^2(L_J)\} - \{E(\rho_J)E(L_J)\}^2
 \end{aligned} \quad (15)$$

Table 1. Possible explanations for the catastrophe insurance puzzle discussed in the literature ¹.

Study	Possible Explanations
(Slovic et al. 1977)	– experimental evidence: people refuse to protect against rare losses with probability below a threshold – convex utility function for losses
(Kunreuther and Slovic 1978)	– consumer disinterest – insurance as investment (only mentioned)
(Kunreuther 1984)	– insurers may not market catastrophe insurance as there may not be enough reinsurance capacity to absorb the probable maximum loss
(Theil 2000)	– insurance decisions in previous experiments may have been affected by the specific format of probabilities, losses, and premiums
(Kunreuther et al. 2001)	– individuals may fail to distinguish between low-probability and zero-probability events
(Kunreuther and Pauly 2004)	– individuals have difficulty processing small probabilities – salience bias
(Akter et al. 2008)	– limited financial resources
(Antwi-Boasiako 2014)	– adverse selection and moral hazard
(Pothon et al. 2019)	– catastrophe premium may be unaffordable for most people
(Lin 2020)	– adverse selection

¹ We would like to thank an anonymous referee for suggesting this table.

The value of Equation (15) is calculated on the basis of the following considerations.

- (1) For the first component of the *I* variance term, the value of $1/\rho_I = 20$ was used for the low-probability event in Section 3.2, implying $\rho_I = 0.05$. For a binary random variable, one has $Var(\rho_I) = \rho_I(1 - \rho_I) = 0.05 \cdot 0.95 = 0.0475$. Therefore, with $E^2(\rho_I) = 0.05^2 = 0.0025$, the bracket amounts to 0.045.
On the basis of the losses caused by ten major hurricanes¹⁰ and assuming equal probabilities of their occurrence, one obtains $Var(L_I) = 449 \cdot 10^6$ and $E^2(L_I) = 1240^2 = 1.54 \cdot 10^6$, resulting in a value of $447.46 \cdot 10^6$ for the second bracket. After multiplication by 0.045, one obtains $20.14 \cdot 10^6$ for the product.
- (2) $\{E(\rho_I)E(L_I)\} = (0.05 \cdot 1240)^2 = 2460$, the *I* component of Equation (15) amounts to $20.14 \cdot 10^6$.
- (3) For the high-probability *J* risk associated with the cell phone, $1/\rho_J = 5$ as stated in Section 3.2. Therefore, $Var(\rho_J) = \rho_J(1 - \rho_J) = 0.2 \cdot 0.98 = 0.196$ and $E^2(\rho_J) = 0.2^2 = 0.04$, so the first bracket equals 0.192. For the second bracket, the average sales price of a cell phone in 2023 was USD 790.¹¹ On the basis of the 18 quotes listed there, the variance is $Var(L_J) = 116,843$; therefore, the value of the second bracket amounts to $116,843 - (0.2 \cdot 970)^2 = 37,636$, resulting in $0.196 \cdot 37,636 = 7377$ for the product. After deduction of $\{E(\rho_J)E(L_J)\} = (0.2 \cdot 0.2 \cdot 970)^2 = 1505$, one obtains 5872 for the *J* component of Equation (9). In all, $Var(\rho_I L_I) + Var(\rho_J L_J)$ amounts to $20.14 \cdot 10^6 + 5872 = 20.146 \cdot 10^6$. Multiplied by $\frac{1}{2}$ and $R_R = 2.5$ and divided by $W = 416,000$, the maximum WTP for certainty is an estimated $0.0605 \cdot 10^3 = \text{US\$ } 60.5 \text{ p.a.}$

Conclusion 4: From a risk-averse individual’s perspective, the combination of a low-probability–high-consequence and a high-probability–low-consequence insurance contract constitutes a good investment as indicated by the combined risk premium. Application of the Arrow-Pratt (1964) formula to the sum of the variances of catastrophe and cell phone insurance results in an estimated maximum willingness-to-pay of USD 60 p.a. as of 2020.

Therefore, the starting point of this analysis can be confirmed: the intuition that combining coverage of a low-probability–high-consequence with a high-probability–low-

consequence event may be a worthwhile investment for a risk-averse decision maker and may therefore alleviate the catastrophe insurance puzzle.

6. Concluding Remarks

The consistent missing demand for catastrophe insurance and other low-probability–high-consequence risks is often referred to as the catastrophe insurance puzzle (CIP). People show reluctance to insure low-probability events, and this behavior is particularly pronounced for some disasters. Individuals rather prefer to insure against high-probability–low-consequence events.

According to Kunreuther and Slovic (1978), one driver of this behavior could be that people think of insurance as an investment. Interestingly, this early explanation from 1978 has never been studied in detail or even been evaluated more carefully in the literature. It was the research question of this article. Using a simple annuity model with fixed annual premiums and expected indemnity payouts, it turns out that even an individual characterized by a degree of risk aversion found in the literature is unlikely to purchase insurance covering a low-probability–high-consequence event. However, combining a low-probability with a high-probability loss event insurance contract may be a way to incentivize individuals to purchase catastrophe risk coverage, alleviating the catastrophe insurance puzzle.

Indeed, insuring against hazards which do not occur in most cases might be a bad investment most of the time. In this paper, this intuition is shown to be correct using a simple model of corporate investment in the absence of risk aversion for both low-probability–high-consequence and high-probability–low-consequence insurance, as well as their combination (Conclusion 1). Only for risk-averse individuals with a high subjective rate of discount is the benefit of certainty sufficiently moved towards the present to make coverage of a low-probability–high-consequence and high-probability–low-consequence event a good investment (Conclusion 2). In contrast, using cell phone insurance as an example, coverage of a high-probability–low-consequence event constitutes a good investment regardless of the degree of risk aversion, given fair premiums (which fully account for the rather high deductibles) (Conclusion 3). These findings go a long way towards explaining the CIP. Finally, demand for low-probability–high-consequence insurance can be enhanced by combining its coverage with coverage of a high-probability–low-consequence event in one policy. Indeed, the (problematic) summation of the risk premia as indicators of willingness-to-pay suggest this constitutes a good investment. When the Arrow-Pratt (1964) formula for deriving maximum willingness-to-pay for certainty is applied to the variances of the product of probability of loss and amount of loss, one arrives at a clearly positive value in the case of catastrophe and cell phone insurance (Conclusion 4).

To alleviate this unfavorable insurance purchase syndrome, combining a low-probability with a high-probability loss insurance contract may be a way to incentivize individuals to purchase catastrophe risk coverage. Depending on the country, such bundling of two very different risks within one single policy might, however, not always be possible due to regulation regarding multi-peril insurance contracts. Furthermore, as outlined by Kunreuther and Michel-Kerjan (2013), it seems worthwhile to evaluate catastrophe risk management strategies involving private–public partnerships to help overcome the lacking demand for low-probability–high-consequence insurance. These partnerships may include multi-year insurance contracts using risk-based premiums coupled with mitigation loans, or vouchers to address affordability for low-income homeowners, tax incentives, improved building codes, and other regulations.

There are a number of limitations to this analysis. First, corporate insurance is typically purchased by managers who may act in their own interest, reflecting risk aversion very similar to the demand for insurance by individuals. Second, the fact that catastrophe insurance imposes deductibles has been neglected throughout on the grounds that they are much smaller relative to the amount of loss than those characteristic of low-consequence (cell phone) insurance. Next, the assumption of fair prices can be justified only in view of its

simplifying the analysis, because loadings tend to be substantial, especially in catastrophe insurance, where losses need to be verified by specialists (this may be changing due to the spread of index insurance and artificial intelligence). Finally, the analysis in this paper uses a constant discount rate; however, behavioral economics and neuroeconomics research have shown deviations from this assumption in favor of what is referred to as hyperbolic discounting. Under this assumption, individuals discount the value of a later payment by a factor increasing in the length of the delay. Under this more realistic depiction of human behavior when discounting the future, the results of this article would be even stronger.

Still, two findings prove to be robust: taking an investment approach regarding the purchase of insurance can explain the catastrophe insurance puzzle, and combining a low-probability with a high-probability insurance contract may well overcome the lacking demand for low-probability–high-consequence insurance indicated by this puzzle.

Author Contributions: Both authors contributed equally to this article. All authors have read and agreed to the published version of the manuscript.

Funding: No funding was received for writing this article.

Data Availability Statement: No data were used for this article.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

- ¹ Interestingly, the terms and conditions included a clause stating that there is no payment in the case of no disaster.
- ² For instance, a purchaser of insurance may feel some regret when the covered loss does not occur even after years of having paid the annual premium. Whether this constitutes a behavioral bias is unclear because it can be derived from rational behavior reflecting a consistent preference structure Eeckhoudt et al. (2018).
- ³ See, for instance, Kunreuther et al. (2001). This argument seems in conflict with Prospect Theory which predicts that very small probabilities are *overestimated* rather than ignored Kahneman and Tversky (1979, 1992). Yet, overestimation does not necessarily entail preventive action (such as buying insurance coverage) if the estimated probability of occurrence is still very low.
- ⁴ See Kanhaiya et al. (2022).
- ⁵ Assume a coin to land head (=a loss) with probability ρ and tail (=no loss) with $(1 - \rho)$. What is E_m , the expected number of tosses until the first head appears? With probability ρ , $m = 1$ and with $(1 - \rho)$, one obtains tail, and a new independent toss is made. Thus, $E_m = \rho + (1 - \rho)(1 + E_m)$ which can be solved for $E_m = 1/\rho$.
- ⁶ See note 5.
- ⁷ See note 5.
- ⁸ See, for instance, DataIntelto (2019).
- ⁹ See, for instance, <https://www.redfin.com/state/Florida/housing-market>, last accessed 1 December 2023.
- ¹⁰ See, for instance, Insurance Information Institute (2023).
- ¹¹ See note 5.

References

- Aditto, Satit. 2011. Risk Analysis of Smallholder Farmers in Central and North-East Thailand. Ph.D. thesis, Lincoln University Canterbury, Lincoln, New Zealand.
- Akter, Sonia, Roy Brouwer, Saria Chowdhury, and Salina Aziz. 2008. Determinants of Participation in a Catastrophe Insurance Programme: Empirical Evidence from a Developing Country, Australian Agricultural and Resource Economics Society 52nd Annual Conference 2008. Available online: <https://ageconsearch.umn.edu/record/5984/> (accessed on 1 December 2023).
- Andor, Mark A., Daniel Osberghaus, and Michael Simora. 2020. Natural Disasters and Governmental Aid: Is there a Charity Hazard? *Ecological Economics* 169: 106534. [CrossRef]
- Antwi-Boasiako, Benjamin A. 2014. Why Do Few Homeowners Insure against Natural Catastrophe Losses? *Review of Economics* 65: 217–40. [CrossRef]
- Arrow-Pratt, John. 1964. Risk aversion in the small and in the large. *Econometrica* 32: 122–36. [CrossRef]
- Brown, Orice W. 2011. Testimony Before the Committee on Banking, Housing, and Urban Affairs, U.S. Senate regarding Flood Insurance. Available online: <https://www.gao.gov/assets/gao-11-670t.pdf> (accessed on 2 December 2023).
- Browne, Mark J., Christian Knoller, and Andreas Richter. 2015. Behavioral bias and the demand for bicycle and flood insurance. *Journal of Risk and Uncertainty* 50: 141–60. [CrossRef]

- DataIntel. 2019. Global Catastrophe Insurance Market, Forecast 2023 to 2031. Available online: <https://dataintel.com/report/catastrophe-insurance-market/> (accessed on 2 December 2023).
- Doherty, Neil A. 1985. *Corporate Risk Management, a Financial Exposition*. New York: McGraw-Hill.
- Eeckhoudt, Louis, Anna Maria Fiori, and Emanuela Rosazza Gianin. 2018. Risk aversion, loss aversion, and the demand for insurance. *Risks* 6: 60. Available online: <https://www.mdpi.com/2227-9091/6/2/60> (accessed on 2 December 2023). [CrossRef]
- Fracasso, Lais, Fernanda Maria Müller, Henrique Pinto Ramos, and Marcelo Brutti Righi. 2023. Is there a risk premium? Evidence from thirteen measures. *Quarterly Review of Economics and Finance* 92: 182–99. [CrossRef]
- Friedman, Bernard. 1974. Risk Aversion and the Consumer Choice of Health Insurance Option. *Review of Economics and Statistics* 56: 209–14. [CrossRef]
- Friend, Irwin, and Marshall E. Blume. 1975. The demand for risky assets. *American Economic Review* 65: 900–22.
- Harrison, Glenn W., Morten I. Lau, and Melonie B. Williams. 2002. Estimating individual discount rates in Denmark: A field experiment. *American Economic Review* 92: 1606–17. [CrossRef]
- Hays, Donald, and Briana Sullivan. 2022. Current Population Reports. The Wealth of Households: 2020. Available online: <https://www.census.gov/content/dam/Census/library/publications/2022/demo/p70br-181.pdf> (accessed on 2 December 2023).
- Insurance Information Institute. 2023. Facts and Statistics: Hurricanes. Available online: <https://www.iii.org/fact-statistic/facts-statistics-hurricanes> (accessed on 2 December 2023).
- Kahneman, Daniel, and Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. In *Handbook of the Fundamentals of Financial Decision Making: Part I*. Singapore: World Scientific, pp. 99–127.
- Kahneman, Daniel, and Amos Tversky. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5: 297–323.
- Kanhaiya, K., R. Pradeep, and K. Vineet. 2022. Mobile Phone Insurance Market Size, Share, Competitive Landscape and Trend Analysis Report by Phone Type, Allied Market Research. Chapter 6.1.1. Market Size. Available online: www.alliedmarketresearch.com/mobile-phone-insurance-market (accessed on 2 December 2023).
- Kunreuther, Howard. 1984. Causes of Underinsurance against Natural Disasters. *Geneva Papers on Risk and Insurance* 31: 206–20. [CrossRef]
- Kunreuther, Howard, and Erwin Michel-Kerjan. 2013. Managing catastrophic risks through redesigned insurance: Challenges and opportunities. In *Handbook of Insurance*. New York: Springer Science and Business Media, pp. 517–46.
- Kunreuther, Howard, and Mark Pauly. 2004. Neglecting disaster: Why don't people insure against large losses? *Journal of Risk and Uncertainty* 28: 5–21. [CrossRef]
- Kunreuther, Howard, and Paul Slovic. 1978. Economics, psychology, and protective behavior. *American Economic Review* 68: 64–69.
- Kunreuther, Howard, Nathan Novemsky, and Daniel Kahneman. 2001. Making low probabilities useful. *Journal of Risk and Uncertainty* 23: 103–20. [CrossRef]
- Laury, Susan K., Melayne M. McInnes, and Todd Swarthout. 2009. Insurance decisions for low-probability losses. *Journal of Risk and Uncertainty* 39: 17–44. [CrossRef]
- Lin, Xiao. 2020. Risk awareness and adverse selection in catastrophe insurance: Evidence from California's residential earthquake insurance market. *Journal of Risk and Uncertainty* 61: 43–65. [CrossRef]
- Meyer, Donald, and Jack Meyer. 2005. Relative risk aversion: What do we know? *Journal of Risk and Uncertainty* 31: 243–62. [CrossRef]
- Moscone, Francesco, and Elisa Tosetti. 2010. Health expenditure and income in the United States. *Health Economics* 19: 1385–403. [CrossRef]
- Pothon, Adrien, Sylvain Buisine, and Pierre-Yves Bard. 2019. California earthquake insurance unpopularity: The issue is the price, not the risk perception. *Natural Hazards and Earth System Sciences* 19: 1909–24. [CrossRef]
- Schlesinger, Harris, and Emilio Venezian. 1986. Insurance markets with loss-prevention activity: Profits, market structure, and consumer welfare. *RAND Journal of Economics* 17: 227–38. [CrossRef]
- Slovic, Paul, Baruch Fischhoff, Sarah Lichtenstein, Bernard Corrigan, and Barbara Combs. 1977. Preference for Insuring Against Probable Small Losses: Insurance Implications. *Journal of Risk and Insurance* 44: 237–58. [CrossRef]
- Stapleton, Richard, and Qi Zeng. 2018. Downside risk aversion and the downside risk premium. *Journal of Risk and Insurance* 85: 379–95. [CrossRef]
- Theil, Michael. 2000. To insure or not to insure? Considerations on irrational strategies to take out insurance. *Geneva Papers on Risk and Insurance-Issues and Practice* 25: 288–95. Available online: <https://link.springer.com/article/10.1111/1468-0440.00066> (accessed on 2 December 2023).
- Zweifel, Peter, Roland Eisen, and David Eckles. 2021. *Insurance Economics*, 2nd ed. Cham: Springer.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

The Impact of ESG Rating on Hedging Downside Risks: Evidence from a Weight-Tilted Hang Seng Index

Joseph K. W. Fung ¹, F. Y. Eric Lam ² and Yiuman Tse ^{3,*}

¹ Lee Shau Kee School of Business and Administration, Hong Kong Metropolitan University, Hong Kong, China; jkwfung@hkmu.edu.hk

² Citigroup, 388 Greenwich Street, New York, NY 10013, USA; fy.eric.lam@citi.com

³ College of Business Administration, University of Missouri, St. Louis, MO 63121, USA

* Correspondence: tseyi@umsl.edu

Abstract: The study examines the return performance and resilience to market volatility of the recently introduced environment, social/sustainable, and governance (ESG) weight-tilted Hang Seng index compared to its parent, the Hang Seng index. The ESG-infused index has a higher mean return and lower return volatility than the parent index, although the differences are statistically and economically insignificant, a result consistent with the high correlation between the two index returns. Most importantly, the ESG weight-tilted index is more resilient to volatility spikes than the parent index and, therefore, has lower downside risks. The overall results show that stocks with high ESG ratings are less susceptible to trading pressures triggered by volatility-induced turnovers. The paper contributes to the literature by providing significant incremental information on the emerging market for ESG-related equity products in Hong Kong.

Keywords: ESG impact investing; index investing; ESG weight-tilted index; downside risk

JEL Classification: F64; G12; G15

Citation: Fung, Joseph K. W., F. Y. Eric Lam, and Yiuman Tse. 2024. The Impact of ESG Rating on Hedging Downside Risks: Evidence from a Weight-Tilted Hang Seng Index. *Journal of Risk and Financial Management* 17: 57. <https://doi.org/10.3390/jrfm17020057>

Academic Editor: Thanasis Stengos

Received: 15 December 2023

Revised: 22 January 2024

Accepted: 29 January 2024

Published: 31 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The environmental, social/sustainable, and governance (ESG) rating and performance of a company affect its market value (Ward and Wu 2019). On the supply side, a company may obtain government subsidies and/or reductions in levies by improving the firm's ESG scores, while on the demand side, stocks with high ESG performance attract ESG-conscious individuals and norm-constrained institutional investors. Moreover, investors whose focus is on risk-adjusted returns can further fuel demand for stocks with high ESG ratings should the scores be priced in the market. All the above factors combine to produce a positive feedback loop, which in effect can create a win-win situation for the high-ESG-rated companies and their shareholders, which include ESG advocates, profit-oriented agents, and ESG norm-constrained and -unconstrained institutional investors. Consequently, firms with high ESG ratings have a lower cost of equity capital and a higher valuation than those at the other end of the measure. The existence of the ESG premium in firm values can propel both financial and real production activities to align economic goals with social welfare objectives.

Zhang et al. (2023) propose three channels through which fund ESG performance could affect fund downside risk: the firm channel in which the risk-mitigation effect of ESG is reflected at the firm level, the diversification channel in which the portfolio concentration of high-ESG funds can aggravate downside risk, and the flow channel in which high-ESG funds can attract investor flows and, consequently, reduce downside risk. The relationship between ESG ratings and downside risk depends on the relative force of these three channels.

We postulate that firms with high ESG ratings are more resilient to market volatility than those with low ESG ratings. Particularly, the preference for high-ESG firms makes them less susceptible to selling pressure in stressful market situations, as the ESG-conscious clientele (investors) are less prone to speculative trading; in addition, the speculative value of such firms is limited since they are already priced at a premium and less attractive to speculators.

Our main contribution is a direct test of the above hypothesis, leveraging the first ESG-enhanced stock market index (the ESG weight-tilted Hang Seng Index HSI_{ESG}) derived directly from the Hang Seng Index (HSI). We examine the performance of the HSI_{ESG} relative to its parent (HSI) under extreme market volatility conditions. A finding that HSI_{ESG} outperforms the parent index under extremely volatile market conditions will indicate the significance of the preference buffer.

The value-weighted free-float-adjusted HSI is the gauge of the blue-chip stocks listed on the Stock Exchange of Hong Kong (SEHK). The index is also the underlying asset of the first Hong Kong index exchange-traded fund (ETF)—that is, the Tracker Fund. The HSI_{ESG} is constructed by shifting the original HSI portfolio weights from the lower to higher ESG-rated firms while maintaining the set of constituent stocks of the parent index. The tilts are based on the ESG scores provided by the Hong Kong Quality Assurance Agency (HKQAA). This new setting that we identify allows a direct test of ESG's marginal impact in terms of the tilted weights' effect on the performance and volatility–return characteristics of the ESG-infused index relative to its parent.

However, unlike the highly distinguishable ESG-driven performance differential of a best-in-class index compared to its broad-based parent index, such as the S&P 500 (Giese et al. 2019a), we expect it would be difficult to discern how the tilted weights affect the performance of the narrow-based HSI because the ESG-infused index and parent index are highly correlated. Moreover, Friede et al. (2015), in their comprehensive meta-analysis, document a weak correlation between ESG and the performance of equities investments, a finding that further lowers the expectation that the HSI_{ESG} can significantly outperform the parent index. Indeed, our study finds that the returns of the weight-tilted HSI_{ESG} and the parent index have a correlation of over 99%, which explains why only minor improvements in the performance metrics (mean and volatility of returns) are found in the weight-tilted index. Nevertheless, the paper shows a subtle but significant negative and asymmetrical relationship between the average returns of the two indexes and the change in index option-implied market volatility. The asymmetrical volatility–return relationship is defined by an observation that the positive return associated with a drop in volatility is less than the magnitude of the negative return triggered by an equal rise in volatility.

We want to determine the performance of the ESG-enhanced index relative to its parent index. If the HSI_{ESG} underperforms the parent index, companies may not pursue ESG practices. The study is guided by (1) what we believe is crucial for driving companies to improve ESG ratings—that is, the return to shareholders (particularly during market downturns) from investing in high ESG firms, and (2) to explore whether the differential asymmetrical volatility–return relationships between the two indexes can explain the difference in the holding period return.

We report the performance comparisons between the HSI_{ESG} and HSI for different volatility conditions. We use both regression and non-parametric analyses to examine the asymmetrical relationships between the index returns and option-implied market volatility. We use the VH_{SI} because the option-implied volatility index is a better measure for ex ante uncertainty than the standard deviation of stock returns. We also analyze the sensitivity of HSI_{ESG} and HSI returns to different systematic risk factors. We find that the returns in both indexes have a negative relationship with changes in the VH_{SI}. Most importantly, this study finds that the response to the volatility of the ESG-infused index return is significantly weaker than that of the parent index during volatile periods, a crucial phenomenon that led to the substantially higher holding period return of the HSI_{ESG} than that of the HSI. This result supports the proposition that firms with high ESG ratings are less susceptible

to trading pressures triggered by volatility-induced turnovers. The evidence also shows that the ESG weight-tilted index is more resilient to volatility spikes than the parent index, indicating that stocks with high ESG ratings can be a hedge against market downside risks. In particular, our results suggest that the firm-level channel and the flow channel are more important than the diversification channel postulated in Zhang et al. (2023). This paper also contributes to the literature by providing significant incremental information on the emerging market for ESG-related equity products.

The rest of the paper is organized as follows. Section 2 reviews related studies. Section 3 describes the data and methodology for the empirical analysis. Section 4 summarizes and interprets the findings. Section 5 concludes the paper.

2. Literature Review

2.1. Evidence of the Market's Preference for ESG-Related Financial Products

The International Monetary Fund (2019) finds that the demand for ESG equity investment funds has accelerated in recent years. Conversely, Brown Brothers Harriman's (2019) survey reveals that ESG ETFs are among the top five ETF sectors that investors prefer to be available in the Hong Kong market. Furthermore, Moody's (2020) study finds that stock indexes attract greater interest when the data compiler incorporates ESG factors in the index products. As ESG-related securities attract fund flows, the funding costs and capital constraints of firms with high ESG ratings can be substantially reduced through the issuance of equity securities. Consequently, the lower required return produces healthier valuations of the stocks and creates higher risk-adjusted returns for investors.

Wu and Juvyns (2020) show that the growth in fund flows into ESG-related equities was uninterrupted by the economic and financial turmoil caused by the COVID-19 pandemic. For instance, in the United States during Q1 2020, ESG-related open-end mutual funds and ETFs received close to USD 10 billion of capital inflows, an amount that is more than half of the total for the full year of 2019. During the same period, the market for ESG ETFs experienced only two weeks of insignificant outflows and the MSCI ESG Leaders Indexes outperformed their extremely volatile market benchmarks (Authers 2020). The above findings show that the prices of ESG-related equity products can weather the downside pressure with the support of norm-constrained institutions and ESG-advocate investors in general.

Giese et al. (2019a) emphasize the benefits to investors in incorporating ESG scores in index construction, as the ESG-tilted index portfolio combines the value of ESG and passive investment in a high-quality well-diversified portfolio. The available ESG-tilted index portfolios can attract large-scale investing toward companies with high ESG ratings, which creates a positive feedback effect, as good ESG practices increase demand for the ESG portfolio and enhance the risk-adjusted return. Conversely, if ESG practices can improve the risk-adjusted return to investors, then ESG-infused financial products and index funds or ETFs also appeal to investors whose main focus is on the potential financial benefits rather than social benefits, an aspect that can further fuel the demand for ESG-related index products.

2.2. Potential Financial Benefits from Investing in Companies with High ESG Ratings

An extensive number of studies have examined the association between ESG ratings and firms' financial performance. Friede et al. (2015) provide a comprehensive meta-analysis that covers over 2200 primary studies and survey articles published over 40 years since 1970. The study shows that over 62% of the primary studies find a positive relationship between ESG rating and corporate financial performance (CFP); the relationships are stable over time and are stronger for emerging markets. The CFP metrics used in the meta-analysis include accounting and market-based risk–return measures.

Gregory et al. (2014) argue that high ESG ratings and performance improve cash flows to shareholders, as ESG attributes strengthen a firm's competitiveness, which raises the company's profitability and dividends. Their argument is consistent with Fatemi et al.'s

(2015) findings that high-ESG firms are more likely than those in the low-ESG group to attract and retain dedicated employees and loyal customers. Dunn et al. (2017) show that the MSCI ESG rating is positively associated with the firm's financial performance but negatively related to its risk. To address the correlation-versus-causality criticism made by Krueger (2015), Giese et al. (2019a) provide an empirical analysis of economic explanations of causality. Pulino et al. (2022) report a positive relationship between ESG disclosure and firm performance (measured by EBIT) for large Italian companies. Wasiuzzaman et al. (2022) suggest that regulators should include cultural dimensions in the development of a single global standard for ESG disclosure. Using data from G20 countries, Bissoondoyal-Bheenick et al. (2023) show that large firms tend to invest more in ESG activities and have better media coverage than small firms. This reduces the information asymmetry of major enterprises regarding ESG investments for their stakeholders.

Eccles et al. (2014) argue that ESG reduces systematic risk, as firms with strong ESG characteristics are less susceptible to market-wide shocks due to improvements in their operational efficiency. Therefore, such companies have lower costs of capital than those with weak ESG performance. Hong and Kacperczyk (2009) and El Ghoul et al. (2011) show that the cost of capital can also be a manifestation of information transparency and that such firms are favored by norm-constrained institutional investors. Godfrey et al. (2009) and Oikonomou et al. (2012) report that ESG reduces financial risk, as a firm with a stronger ESG profile has higher compliance standards and better risk management and is, therefore, less vulnerable to idiosyncratic and operational risks than its counterparts. This allows high-ESG firms to avoid costly lawsuits and settlements. Giese et al. (2019b) also find among MSCI-rated firms that companies with high-MSCI ESG ratings have reduced idiosyncratic risk and an increased buffer against market risk. Lins et al. (2017) show that social responsibility helps firms earn trust and social capital during market downturns; their results are further supported by Jin et al. (2023). See also, for example, Cao et al. (2023) and Li et al. (2023) for evidence from the Chinese stock markets.

Conversely, there are concerns that the inclusion of ESG criteria may reduce returns (see, e.g., Nagy et al. 2016) because the ESG tilts might underweight stocks with high risk-adjusted returns and overweight stocks with low risk-adjusted returns. The matter is serious as it is related to the investment fund manager's fiduciary duty. However, such concern was lessened after the US Labor Department opined that ESG-related investment decisions made by pension plans do not violate the fiduciary duty of the sponsor and added that incorporating ESG ratings can create both social and financial benefits, according to Friede et al. (2015). Nevertheless, there are questions raised as to whether the ESG rating is precise. For example, Berg et al. (2022) show that such ratings provided by the six prominent agencies are dispersed and mainly driven by divergences of scope and measurement methodology in addition to the assessor's overall view of a firm. Furthermore, ESG rating might as well be a surrogate for known return predictors; hence, it does not present new valuable information to investors. For example, Melas et al. (2018) show that ESG ratings have a negative association with the value factor (see, e.g., Fama and French 2015). In a similar vein, Authers (2020) argues that ESG investing could be a watered-down version of growth investing, with certain sectors, such as technology and healthcare, being overweight.

Several recent papers provide different results and perspectives about the fund ESG performance. Zhang et al. (2023) propose three channels through which ESG performance will influence fund downside risk. First, the risk-mitigation effect of individual portfolio firms' good ESG practices suggests a negative relationship between ESG ratings and downside risk (the firm-level channel). Second, the concentration of firms with high ESG ratings (or the elimination of firms with low ESG ratings) in a portfolio will lower the benefit of portfolio diversification, resulting in a positive relationship between ESG ratings and downside risk (the diversification channel). Third, firms with high ESG ratings may attract more investors (particularly long-term investors and during market turmoil) and fund flows, leading to a negative ESG rating–downside risk relationship (the flow channel).

Zhang et al. report that the force of the flow channel dominates the other two channels in the Chinese market and find a positive relationship between fund ESG performance and downside risk.

Davoodi et al. (2024) use advanced machine learning and regime-switching models to construct portfolios. They show that separating a time series into different regimes based on volatility and considering different metrics (in addition to historical return and volatility) can enhance fund performance. Rojo-Suárez and Alonso-Conde (2024) examine whether shifts in investor tastes have led the market portfolio to capture ESG preferences. Their results suggest that “efforts by public authorities to promote improvements in corporate ESG performance translate into lower cost of capital, especially in periods of overall declines in corporate ESG performance”.

Our study examines whether the HSI ESG is more resilient to market volatility than its parent, the HSI. However, unlike a best-in-class index—that is, an index portfolio constructed with a subset of top ESG-rated firms in a broad-based index, such as the S&P 500—it is widely known that a weight-tilted narrow-based index is expected to be highly correlated with the parent index (see, e.g., Giese et al. 2019a, 2019b). Consequently, it is highly unlikely that the HSI ESG can significantly outperform the parent HSI in any aspect. Therefore, the finding of a significant difference in the risk and return profiles between the HSI ESG and its parent HSI would provide a strong testimony that the ESG-tilted weights have a material impact on the index performance and that the ESG-infused portfolio is more resilient to market volatility than the parent index due to a preference buffer. Following the arguments of Zhang et al. (2023), our results suggest that the firm-level channel and the flow channel are more important than the diversification channel.

3. Data and Methodology

3.1. Data

Although ESG investing is new to the Hong Kong equities market, the asset management industry has already begun to internalize the opportunity. The Hang Seng Indexes Co. Ltd. (Hong Kong, China), the provider of the HSI and various major Hong Kong stock market benchmark indexes, launched the HSI ESG on 14 May 2019. The HSI ESG and its parent index are identical in all respects except that the HSI ESG is constructed by shifting the index weights from firms with low ESG ratings to firms with high ESG ratings, where tilts are based on the ESG scores compiled by the HKQAA. The portfolio weight of a single stock has been capped at 8% for the parent index; the ceiling remains in effect for the tilt-adjusted weight of the HSI ESG. Weightings are not fully disclosed by the HSI. There was no index exclusion from the HSI during the period. If a stock is (not) in the HSI, it is (not) in the HSI ESG—survivorship should not affect the comparison made in the study. Furthermore, both indexes are subject to quarterly review.

The SEHK has required all listed companies to provide an annual ESG report to the public since 2016, which agrees with the Special Administrative Region (SAR) government’s vision to become China’s international green and sustainable financial hub. These conditions facilitate the index provider to launch the ESG weight-tilted index to meet the potential market demand.

The index provider has backdated the HSI ESG to 8 September 2014. The overall sample covers 1751 daily observations for the period 8 September 2014–31 October 2021. The availability of the backdated sample allows a comparison of the findings between the prelaunch period (8 September 2014–13 May 2019; $N = 1148$) and the postlaunch period (14 May 2019–31 October 2021; $N = 603$). However, it is expected that as the ESG score data were also available during the prelaunch period, the market at large should have included the information in their index portfolio, and it is expected that the key results from the two subperiods are similar according to Friede et al.’s (2015) finding that the correlation between ESG and CFP is stable over time. Our paper uses daily data of the HSI ESG, HSI, and VHSI retrieved from the Bloomberg terminal; the daily market factors for the Fama and Macbeth regression analysis are obtained from the Kenneth French library. We use

daily data to capture the time-series dynamics of the index return in Hong Kong dollars. However, the main results (available upon request) are qualitatively the same using weekly and monthly returns.

3.2. Methodology

3.2.1. Performance Measurements and Comparisons

Conventional risk and return and other performance measures including the distributions of return, information ratio (IR), Sortino ratio, value at risk (VaR), expected shortfall (CVaR), and maximum drawdown are used to compare the risk and return between the ESG-infused index and the parent indexes for the overall period and between the prelaunch and postlaunch subperiods. The IR is r/σ , where r is the mean daily return and σ is the standard deviation of daily returns. The Sortino ratio is r/σ_d , where σ_d is the standard deviation of negative returns (downside deviation). VaR measures the greatest possible losses over a specific period at a particular percentile of return. Expected shortfall (also called conditional VaR) is the expected loss during the period, conditional on a loss greater than the particular percentile of the loss distribution. Maximum drawdown calculates the downside risk as the difference between the peak and trough index values in the percentage of the peak index value.

3.2.2. Tests for the Difference in Asymmetrical Volatility–Return Relationships between the HSIESG and HSI

We use the following conventional multiple regressions to examine the volatility–return relationships for both indexes and across the two subperiods. We examine the relationship of index returns ($\Delta \ln HSI_t$) with the change ($\Delta VHSl_t$ in Equation (1)) and percentage change ($\Delta \ln VHSl_t$ in Equation (2)) in the option-implied volatility. Equations (1) and (2) are as follows:

$$\Delta \ln HSI_t = \alpha_0 + \alpha_1 \Delta VHSl_t \alpha_2 D_t \times \Delta VHSl_t + \alpha_3 D_t + e_t \quad (1)$$

$$\Delta \ln HSI_t = \beta_0 + \beta_1 \Delta VHSl_t + \beta_2 D_t \times \Delta \ln VHSl_t + \beta_3 D_t + e_t, \quad (2)$$

where $\Delta \ln HSI_t = \ln HSI_t - \ln HSI_{t-1}$. D_t is a dummy variable with $D_t = 1$ if $\Delta VHSl_t < 0$ and 0 otherwise. α_1 and β_1 are expected to be negative. A negative α_2 or β_2 indicates an asymmetrical volatility–return relationship and that the negative market response to an increase in volatility is stronger than a decrease in volatility. We replicate the regressions with the HSIESG.

All time-series regressions use the heteroskedasticity and autocorrelation (HAC) Newey–West standard errors to account for both heteroskedasticity and autocorrelation in the error term. We use 10 lags in the Newey–West estimation, and the results are similar using different lags. Next, we group the returns of the 2 indexes within each of the 10 bins defined by the deciles of the rate of implied volatility change ($\Delta \ln VHSl$). Decile 1 (the bottom volatility change bin) contains the index returns on days with the sharpest drop in market volatility, while decile 10 (the top volatility change bin) includes the index returns on days with the steepest rise in market volatility.

We use the HAC standard errors of Newey and West (1984) to test the statistical significance of the difference of mean returns. Furthermore, we conduct a robustness check on the statistical significance of the mean returns and their differences using a bootstrapping method by resampling the returns 10,000 times to avoid the problem associated with the non-normality of the return distribution. For a return in a volatility change decile, from all the resampled means, we examine the observed distribution of the mean to see if the percentile range [0.5, 99.5] contains the value 0. If not, we define the mean of the return distribution to be 1% statistical significance. We report the results of the robustness tests in Appendix A.

3.2.3. Tests for Differential Exposure to Various Investment Factors

Fama and French’s (2015) five-factor models with a momentum factor are used in this paper to assess the differential exposure to various market factors between the ESG weight-tilted HSI_{ESG} and the parent index. We run the following factor model:

$$\Delta \ln HSI_t = \gamma_0 + \gamma_1(Mkt-RF)_t + \gamma_2SMB_t + \gamma_3HML_t + \gamma_4RMW_t + \gamma_5CMA_t + \gamma_6WML_t + e_t, \quad (3)$$

where *Mkt-RF* is the market risk premium for developed international markets. *SMB* (size factor) is the return differential between small- and large-cap stocks. *HML* (P/B factor) is the return differential between high- and low-book-to-market stocks. *RMW* (profitability factor) is the return differential between stocks of high and low operating profitability. *CMA* (pro-growth factor) is the return differential between aggressive and conservative companies. *WML* (momentum factor) is the return differential between winners and losers. We replicate the regression with the HSI_{ESG}.

The results allow an examination of the relative performance between the ESG-infused index and the parent index. In their comparative study, Nagy et al. (2016) show that an ESG-tilt investment strategy that overweighs stocks with higher ESG ratings based on global MSCI data outperforms the benchmark. Finding similar results in Hong Kong would suggest that the ESG tilts have effectively incorporated and reflected the market’s relative preference for firms with high ESG scores.

We can provide more economic significance by adding suitable proxy variables for the ESG-individuals’ holdings of ESG stocks and norm-constrained institutional investors’ holdings of ESG stocks in the prelaunch and postlaunch periods to test the main hypothesis that high-ranked ESG stocks’ performance is resilient to market volatility. We reserve this issue for future research when data are available.

4. Empirical Results and Interpretations of the HSI and HSI_{ESG}

Table 1 shows the summary statistics of daily returns on the ESG weight-tilted HSI_{ESG}, the parent HSI, and daily closing levels of and returns on the option-implied volatility index derived from options written on the HSI. The mean and standard deviation of daily returns of the two indexes are not statistically significantly different for the overall period and the two subperiods.

Table 1. Summary statistics of daily returns of the ESG weight-tilted Hang Seng index (HSI_{ESG}), the parent Hang Seng index (HSI), and daily closing levels of and returns of the Hang Seng option-implied volatility index (VHSI).

	HSI _{ESG} Return	HSI Return	VHSI Close	VHSI Return
Full sample (8 September 2014–October 2021; <i>N</i> = 1751)				
Mean	0.0001	0.0001	20.0736	0.0022
Std dev	0.0115	0.0119	5.6591	0.0655
Median	0.0004	0.0006	19.1400	−0.0072
Max	0.0437	0.0505	64.8000	0.5839
Prelaunch period (8 September 2014–13 May 2019; <i>N</i> = 1148)				
Mean	0.0002	0.0002	19.1027	0.0021
Std dev	0.0109	0.0112	4.8960	0.0624
Median	0.0005	0.0007	18.3200	−0.0058
Max	0.0409	0.0421	41.0100	0.5839
Postlaunch period (14 May 2019–October 2021; <i>N</i> = 603)				
Mean	−0.0001	−0.0001	21.9220	0.0023
Std dev	0.0125	0.0131	6.4966	0.0711
Median	0.0003	0.0003	20.4900	−0.0085
Max	0.0437	0.0505	64.8000	0.5102

The mean daily returns and the standard deviation of the daily returns of the two indexes are not statistically different for the overall sample and the two subperiods. Statistical tests (not reported here) reject the null hypothesis that the return distributions are normal.

Table 2 shows the correlations among daily returns of the HSI ESG and HSI, and the levels and returns of the VHSI. All correlation coefficients are statistically significant at the 1% level. The high return correlation (>99%) between the two indexes suggests that finding any significant differences in the risk and return profiles between the weight-tilted index and the parent index is highly unlikely. The negative and over 60% correlation between the two measures of volatility change and index returns are consistent with the widely documented negative volatility–return relationship in the equity markets. The negative volatility–return relationships have strengthened in the recent period for both indexes, and the change in correlations is qualitatively identical for both indexes. It is useful to mention here that the subperiod results do not reveal significantly different correlation patterns between the prelaunch and postlaunch periods.

Table 2. Correlations among daily returns of the HSI ESG and HSI and levels and returns of the VHSI.

	$\Delta \ln \text{HSI ESG}$	$\Delta \ln \text{HSI}$	VHSI	ΔVHSI
Full sample				
$\Delta \ln \text{HSI}$	0.9933			
VHSI	−0.1710	−0.1656		
ΔVHSI	−0.6507	−0.6609	0.1412	
$\Delta \ln \text{VHSI}$	−0.6414	−0.6534	0.1347	0.9435
Before ESG index launch				
$\Delta \ln \text{HSI}$	0.9956			
VHSI	−0.1633	−0.1619		
ΔVHSI	−0.6258	−0.6351	0.1368	
$\Delta \ln \text{VHSI}$	−0.5941	−0.6049	0.1338	0.9650
After ESG index launch				
$\Delta \ln \text{HSI}$	0.9902			
VHSI	−0.1872	−0.1767		
ΔVHSI	−0.6911	−0.6998	0.1548	
$\Delta \ln \text{VHSI}$	−0.7110	−0.7235	0.1470	0.9368

$\Delta \ln \text{HSI ESG}$ is the continuous compounded daily returns of the HSI ESG, $\Delta \ln \text{HSI}$ is the continuous compounded daily returns of the HSI, VHSI is the HSI option-implied volatility index, ΔVHSI is the daily change in the level of the VHSI, and $\Delta \ln \text{VHSI}$ is the continuous compounded daily returns of the VHSI. All correlation coefficients are statistically significant at the 1% level; the high return correlation (>99%) between the two indexes suggests that it is highly unlikely to find any significant differences in the risk and return profiles between the weight-tilted index and the parent index. The negative and over 60% correlation between the two measures of volatility change and index returns are consistent with the widely documented negative volatility–return relationship in the equity markets. The negative volatility–return relationships have strengthened in the recent period for both indexes, and the change in the correlation patterns are qualitatively similar for both indexes.

Table 3 summarizes the key performance and risk metrics between the HSI ESG and HSI. In general, the HSI ESG has a higher return and a lower return standard deviation than those of the HSI, but, again, the differences are not statistically and economically significant. Specifically, the average (standard deviation of) daily returns are 0.0088% (0.0115) and 0.0084% (0.0119) for the HSI ESG and HSI, respectively. The two subperiod results are qualitatively similar to those of the overall period. However, despite the insignificant differences in the arithmetic mean returns and return standard deviations between the two indexes, the holding period return of the HSI ESG is substantially higher than its parent index by over 67% (i.e., 3.9657% vs. 2.369%), an important result that we will further explore in the subsequent sections.

Table 4 summarizes the test results on the negative and asymmetrical relationships between the index returns and the change in option-implied market volatility. The regression results in Panel A show a highly significant negative relationship between index returns and the two measures in volatility changes. The slope coefficients for both measures of volatility change are negative at the 1% significance level, but the intercepts are mostly insignificant. Panel B shows that the intercept dummy and the slope coefficients are significantly negative at the 1% significance level concerning volatility change (ΔVHSI), an indication of the asymmetrical volatility–return relationship between the returns of the

two indexes and the volatility change. Conversely, the slope coefficient has fully captured the volatility–return relationship for both indexes concerning volatility return ($\Delta \ln VHSI$). The regression test results further confirm the asymmetrical negative volatility–return relationship concerning the raw change in volatility. However, concerning the rate of volatility change, we find a highly significant negative volatility–return relationship in the slope coefficient but not in the intercept term. Furthermore, the subperiod results show no significant difference in the volatility–return relationship between the prelaunch and postlaunch periods.

Table 3. Performance comparisons between the HSI ESG and HSI.

	Overall		Prelaunch Period		Postlaunch Period	
	8 September 2014–19 October 2021; N = 1751		8 September 2014–13 May 2019; N = 1148		14 May 2019–Oct 2021; N = 603	
	HSIESG	HSI	HSIESG	HSI	HSIESG	HSI
Panel A: Performance indicators						
Average daily return (%)	0.0088	0.0084	0.0180	0.0172	−0.0066	−0.0058
Std. dev. of daily returns (%)	0.0115	0.0119	0.0109	0.0112	0.0125	0.0131
Annualized return with daily compounding (%)	2.2448	2.1414	4.6438	4.4228	−1.6464	−1.4485
Holding period return (%)	3.9657	2.3690	14.8513	13.3376	−8.3050	−8.3024
Annualized information ratio mean/std.	0.1234	0.1136	0.2687	0.2493	−0.0831	−0.0697
Annualized Sortino ratio (mean/−ve std.)	0.1702	0.1567	0.3753	0.3481	−0.1133	−0.0954
Panel B: Risk indicators						
Annualized total risk (%)	18.1961	18.8514	17.2829	17.7394	19.8138	20.7955
Annualized downside deviation (%)	13.1882	13.6622	12.3727	12.7052	14.5294	15.1898
Skewness	−0.3325	−0.3166	−0.3282	−0.3172	−0.3270	−0.3032
Kurtosis	2.2559	2.0447	2.2396	2.0758	2.0813	1.7750
VaR @ 95% (%)	−1.905	−2.0266	−1.7577	−1.8222	−2.0559	−2.1933
VaR @ 99% (%)	−3.2057	−3.1761	−2.7991	−2.9723	−3.4547	−4.0334
Expected shortfall (CVaR) @ 95% (%)	−2.7129	−2.8108	−2.5386	−2.6099	−2.9416	−3.0692
Expected shortfall (CVaR) @ 99% (%)	−4.0797	−4.2028	−3.7132	−3.7842	−4.2434	−4.4250
Maximum drawdown (%)	35.9013	35.5914	35.6387	35.5914	26.8623	25.3310

Panels A and B show that the HSIESG has a higher return and lower return standard deviation than the parent index; however, the differences are not statistically and economically significant. Conversely, despite the minor differences in the mean and standard deviation of the daily returns between the two indexes, the HSIESG has an over 67% higher return for the overall holding period compared to the HSI (i.e., 3.9657% vs. 2.369%).

Table 5 summarizes the mean daily returns for the HSIESG and HSI within each of the 10 bins defined by the deciles of the rate of implied volatility change. Decile 1 (the bottom volatility change bin) shows the mean index returns on days with the steepest drop in market volatility, while decile 10 (the top volatility change bin) shows the mean index returns on days with the sharpest rise in market volatility. Consistent with the regression results, the mean returns for both indexes are significantly positive in the bins with an average negative change in volatility and vice versa. The above findings are qualitatively similar for the two subperiods. We use the HAC standard errors of Newey and West (1984) to test the statistical significance of the results. Table A1 in Appendix A reports our robustness test via a bootstrapping method. The results from both tests are qualitatively identical. Most importantly, the asymmetrical response to the volatility of the ESG-infused index is significantly weaker than the parent index. For the overall period, the HSIESG has a lower mean return than the parent index (1.0496% vs. 1.1261%) for days with the highest drops in volatility. The opposite is true for days with the greatest spikes in volatility, where the HSIESG has a less negative mean return than the parent index (−1.6255% vs. −1.7235%).

Table 4. Tests of the negative and asymmetrical relationships between the index returns (HSIESG and HSIESG returns) and change in option-implied market volatility as measured by $\Delta VHSI$ and $\Delta \ln VHSI$ for the overall sample period ($N = 1751$). Dependent variable: HSIESG or HSIESG returns.

Independent Variables	Estimates			
	HSIESG	HSI	HSIESG	HSI
Panel A:				
Intercept	0.0001 *	0.0001	0.0001	0.0001
$\Delta VHSI$	−0.0046 ***	−0.0049 ***		
$\Delta \ln VHSI$			−0.1189 ***	−0.1255 ***
F-value	90.58	86.32	94.81	98.30
R ²	0.423	0.437	0.411	0.427
R ² -adjusted	0.423	0.436	0.411	0.427
Panel B: ($N = 1751$)				
Intercept	0.0017 ***	0.0018 ***	0.0003	0.0003
$\Delta VHSI$	−0.0036 ***	−0.0039 ***		
$D \times \Delta VHSI$	−0.0009 **	−0.0007 *		
$\Delta \ln VHSI$			−0.1166 ***	−0.1249 ***
$D \times \Delta \ln VHSI$			0.0009	0.0052
D	−0.0027 ***	−0.0031 ***	−0.0005	−0.0007
F-value	97.16	92.70	79.81	78.49
R ²	0.434	0.448	0.412	0.428
R ² -adjusted	0.433	0.447	0.411	0.427

Statistical significance levels of 1%, 5%, and 10% are represented by ***, **, and *, respectively. Panel A shows the generic result that there is a significant negative relationship between index returns and change in option-implied market volatility; the slope coefficients for both measures of volatility change are significantly negative at the 1% level. Panel B shows the test results of the asymmetrical impact of volatility change on market returns using a dummy variable, where $D1 = 1$ if $\Delta VHSI > 0$ and 0 otherwise. An asymmetrical impact is observed if positive changes in option-implied market volatility have a stronger impact on the index returns than negative implied volatility changes. The regression results provide strong empirical evidence of a negative relationship between change in option-implied volatility and market returns. The negative slope coefficients are significant at the 1% level, and the results are robust concerning both indexes and different measures of volatility change. A highly significant (at the 1% level) asymmetrical impact of volatility change ($\Delta VHSI$) on returns is observed for both indexes from the coefficient for the intercept dummy, while the asymmetrical impact on the slope coefficients is weaker but still significant at the 5% and 10% levels for the HSIESG and HSI, respectively. Conversely, the asymmetrical effects on either the slope or the intercept term are insignificant concerning the rate of volatility change (i.e., $\Delta \ln VHSI$). All of the above results are valid and essentially identical for the two subperiods (i.e., pre- and postlaunch periods), indicating the absence of effect on the market from the official launch of the HSIESG. The results for the two subperiods are qualitatively similar to those from the overall period. They are not reported here to conserve space but are available upon request.

Table 6 shows that the above-mentioned mean return differentials particularly for the top and bottom decile bins are statistically significant at the 1% level. Moreover, the results from the two subperiods are consistent with those found in the overall period. We also consider whether the market cap is the main driver of the differential asymmetrical risk–return relationship and the significant holding return gap between the two indexes. By examining a sample of snapshots (since the portfolio weights are changing over time) of the two sets of portfolio weights, we find that the tilted weights have generally migrated downward for the largest stocks. The reason is that because the weights of the largest stocks have already reached the cap rate in the parent index, the ESG tilts can shift their weights only downward. Hence, it is unclear whether the ESG performance is related to the market cap of the largest stocks. The table also shows the difference between the mean returns between the HSIESG and HSI ($\Delta \ln HSIESG - \Delta \ln HSI$) within each bin defined according to the volatility change decile. We test the differences using a *t*-test with the HAC standard errors of Newey and West (1984). The results show that the parent index HSI generally outperforms the ESG-infused HSIESG for days with the steepest volatility drop (within the volatility change decile 1 bin), while the opposite is true for days with the sharpest rise in volatility (within the volatility change decile 10 bin). The above findings are similar

for the two subperiods. Table A2 in Appendix A reports the robustness test results via a bootstrapping method. The results from both tests are qualitatively the same.

Table 5. The mean returns of the HSI/ESG ($\Delta \ln \text{HSI/ESG}$) and HSI ($\Delta \ln \text{HSI}$) for each decile of option-implied volatility return ($\Delta \ln \text{VHSI}$) classification for the overall period and the prelaunch and postlaunch subperiods.

Decile	Overall			Prelaunch			Postlaunch		
	$\Delta \ln \text{VHSI}$	$\Delta \ln \text{HSI/ESG}$	$\Delta \ln \text{HSI}$	$\Delta \ln \text{VHSI}$	$\Delta \ln \text{HSI/ESG}$	$\Delta \ln \text{HSI}$	$\Delta \ln \text{VHSI}$	$\Delta \ln \text{HSI/ESG}$	$\Delta \ln \text{HSI}$
1	−9.0953	1.0496 ***	1.1261 ***	−8.5521	0.9139 ***	0.9788 ***	−10.0522	1.3110 ***	1.4060 ***
2	−4.8753	0.6018 ***	0.6380 ***	−4.7405	0.5751 ***	0.5965 ***	−5.1077	0.6366 ***	0.6991 ***
3	−3.3732	0.4504 ***	0.4737 ***	−3.3077	0.4153 ***	0.4319 ***	−3.4804	0.5012 ***	0.5462 ***
4	−2.3367	0.3418 ***	0.3534 ***	−2.2652	0.2854 ***	0.2977 ***	−2.4734	0.4751 ***	0.4931 ***
5	−1.2689	0.1447 ***	0.1519 ***	−1.1902	0.2009 ***	0.2002 ***	−1.4111	0.0582	0.0628
6	−0.1699	0.0390	0.0515	−0.0936	0.0610	0.0663	−0.3035	0.0148	0.0408
7	1.0704	−0.0478	−0.0637 **	1.1353	−0.0483	−0.0587 *	0.9268	−0.0775 **	−0.1038 ***
8	2.6172	−0.2573 ***	−0.2899 ***	2.6346	−0.1870 **	−0.2040 ***	2.5615	−0.3746 ***	−0.4274 ***
9	4.9207	−0.6085 ***	−0.6334 ***	4.7841	−0.5562 ***	−0.5867 ***	5.2033	−0.8251 ***	−0.8652 ***
10	12.7039	−1.6255 ***	−1.7235 ***	11.9064	−1.4779 ***	−1.5480 ***	14.0373	−1.7779 ***	−1.9015 ***

Daily index returns are grouped into 10 bins according to market volatility returns. The decile 1 bin contains the returns of the indexes on days with the steepest drop in market volatility, while the decile 10 bin contains the returns on days with the sharpest rise in market volatility. The mean returns of the indexes are positive on days in the bins with a negative mean volatility return (i.e., deciles 1–6), while the mean returns of both indexes are positive on days in the bins with a positive mean volatility return. The results are consistent for the three periods. The 1%, 5%, and 10% statistical significance levels, indicated by ***, **, and *, respectively, are determined by the HAC standard errors of Newey and West (1984). The positive mean returns for both indexes in the decile 10 bins are less than the magnitudes of the negative returns in the corresponding decile 1 bins, a result consistent for the overall period and the two subperiods. The findings show that an up jump in volatility has a greater impact on the index return than a down jump in volatility. Most importantly, the ESG-infused index has a significantly less asymmetrical volatility–return relationship than the parent index. For the overall period, the HSI/ESG has a lower mean return than the parent index (1.0496% vs. 1.1261%) for days with the highest drops in volatility; the opposite is true for days with the greatest spikes in volatility, and the HSI/ESG has a less negative mean return than the parent index (−1.6255% vs. −1.7235%).

Results reported in Table 6 also shed some light as to why the ESG-infused index has a higher return than the parent index. Negative returns have a greater impact than positive returns on the holding period return. As noted in Table 6, we find that the ESG-infused index has a significantly less negative mean return than the parent index on days with the highest volatility spikes (i.e., the decile 10 bin) while the positive mean return differential ($\Delta \ln \text{HSI/ESG} - \Delta \ln \text{HSI} > 0$) in bin 10 is greater than the magnitude of the negative mean return differential ($\Delta \ln \text{HSI/ESG} - \Delta \ln \text{HSI} < 0$) in bin 1. Hence, the higher holding period return of the ESG-infused parent index is a result of the observed asymmetrical differential response to volatility shocks between the two indexes.

As noted in Table 3, the less negative mean return of the ESG-infused index on days with the sharpest rise in volatility produces a substantially higher holding period return than the parent index (i.e., 3.966% vs. 2.369%) despite the seemingly minor and statistically insignificant difference in the standard deviation of daily returns (i.e., 0.0115% vs. 0.0119%). Although the maximum drawdown of the ESG-tilted index is slightly higher than the parent index by 31 basis points for the overall period, the 95% VaR and the CVaRs of the HSI/ESG for both confidence intervals are lower than those of the parent index. Moreover, the holding period return difference is mainly attributed to the differences in mean returns between the two indexes on days with the greatest drop and the sharpest rise in market volatility (see the interpretation of results for Tables 5 and 6).

Table 6. The differential mean returns between the HSI ESG and HSI ($\Delta \ln \text{HSI ESG} - \Delta \ln \text{HSI}$) for each decile of volatility return classification.

Decile	Overall			Prelaunch	Postlaunch
	$\Delta \ln \text{VHSI}$	$\Delta \ln \text{HSI ESG} - \Delta \ln \text{HSI}$	$\Delta \ln \text{VHSI}$	$\Delta \ln \text{HSI ESG} - \Delta \ln \text{HSI}$	$\Delta \ln \text{HSI ESG} - \Delta \ln \text{HSI}$
1	−9.0953	−0.0765 ***	−8.5521	−0.0649 ***	−10.0522
2	−4.8753	−0.0362 ***	−4.7405	−0.0214 **	−5.1077
3	−3.3732	−0.0233	−3.3077	−0.0166 **	−3.4804
4	−2.3367	−0.0116 ***	−2.2652	−0.0123 **	−2.4734
5	−1.2689	−0.0072	−1.1902	0.0007	−1.4111
6	−0.1699	−0.0125 *	−0.0936	−0.0053	−0.3035
7	1.0704	0.0159 **	1.1353	0.0104 **	0.9268
8	2.6172	0.0326 **	2.6346	0.017 ***	2.5615
9	4.9207	0.0249 **	4.7841	0.0305 ***	5.2033
10	12.7039	0.098 ***	11.9064	0.0701 ***	14.0373

The 1%, 5%, and 10% statistical significance levels, indicated by ***, **, and *, respectively, are determined by the HAC standard errors of Newey and West (1984). The results show that the HSI generally outperforms the HSI ESG for days with the steepest volatility drop, as shown in the results from the decile 1 bin, while the opposite is true for days with the sharpest rise in volatility, as shown in the results from the decile 10 bin. The above findings are similar for the two subperiods. As noted in Table 3, the less negative mean return of the ESG-infused index on days with the greatest rise in volatility produces a substantially higher holding period return than the parent index (i.e., 3.9657% vs. 2.369%) despite the seemingly minor and statistically insignificant difference in the standard deviation of daily returns (i.e., 0.0115% vs. 0.0119%). This result leads to the conclusion that the HSI ESG has a substantially higher holding period return than the parent index because the ESG-infused index has a significantly less negative mean return than the parent index on days with the highest volatility spikes (i.e., the decile 10 bin). Moreover, the positive mean return differential ($\Delta \ln \text{HSI ESG} - \Delta \ln \text{HSI} > 0$) in bin 10 is greater than the magnitude of the negative mean return differential ($\Delta \ln \text{HSI ESG} - \Delta \ln \text{HSI} < 0$) in bin 1. To understand the large difference in the overall holding period return between the two indexes, we calculate the cumulative returns for days included in volatility change in decile 1, decile 10, and the rest of the sample period. We find the following: (1) the cumulative return for days in decile 1 (the top 10% volatility change) is −380% and −399% for the HSI ESG and HSI, respectively; (2) the cumulative return for days in decile 10 (the bottom 10% volatility change) is −343% and −355% for the HSI ESG and HSI, respectively; (3) the cumulative returns for all other days (deciles 2–9, both deciles included) are 41.74% and 46.01% for the HSI ESG and HSI, respectively.

To understand the large difference in the overall holding period return between the two indexes, we calculate the cumulative returns for days included in volatility change decile 1, decile 10, and the rest of the sample period. We find the following: (1) the cumulative return for days in decile 1 (the top 10% volatility change) is −380% and −399% for the HSI ESG and HSI, respectively; (2) the cumulative return for days in decile 10 (the bottom 10% volatility change) is 343% and 355% for the HSI ESG and HSI, respectively; (3) the cumulative returns for all other days (deciles 2–9, both deciles included) are 41.74% and 46.01% for the HSI ESG and HSI, respectively. This result leads to the conclusion that the HSI ESG has a substantially higher holding period return than the parent index because the ESG-infused index has significantly less negative returns than the parent index during days with the highest volatility spikes.

Tables 7 and 8 summarize the test results on the sensitivity of the HSI ESG and HSI returns to systematic risk measures via the Fama–French multifactor capital asset pricing model. Table 7 summarizes the results using market factors of the developed international markets, while Table 8 shows the results using the systematic risk factors of the Asia–Pacific markets excluding Japan. Both factors of SMB and HML for developed international markets (Table 7) and Asia–Pacific markets excluding Japan (Table 8) are used. The analysis is extended to examine whether the official launch (vis-a-vis the prelaunch period) of the ESG-tilted weights has a material impact on the performance of the HSI ESG compared to the parent index. The overall results of Tables 7 and 8 show that the HSI and HSI ESG have similar sensitivities to the Fama–French five factors and the momentum factor, indicating that our results are not driven by the different sensitivities of these two indexes to the conventional factors. The results from the two subperiods are similar to those for the overall

sample period. The two subperiods are the prelaunch and postlaunch periods as shown in Figure 1.

Table 7. Sensitivity of HSIESG and HSI returns to systematic risk factors in developed international markets.

	HSIESG	HSI	HSIESG	HSI	HSIESG	HSI	HSIESG	HSI	HSIESG	HSI
Full Sample Period										
Intercept	−0.0002	−0.0002	−0.0002	−0.0003	−0.0002	−0.0003	−0.0003	−0.0003	−0.0003	−0.0003
Mkt-RF	0.5276 ***	0.5473 ***	0.6689 ***	0.6994 ***	0.6682 ***	0.6986 ***	0.6083 ***	0.6310 ***	0.6085 ***	0.6312 ***
SMB			0.7783 ***	0.8310 ***	0.7790 ***	0.8318 ***	0.7430 ***	0.7895 ***	0.7427 ***	0.7891 ***
HML			0.0007	−0.0199	−0.0204	−0.0430	0.3829 ***	0.4000 ***	0.4027 ***	0.4206 ***
RMW							0.1795	0.1759	0.1858	0.1825
CMA							−0.8075 ***	−0.8954 ***	−0.8148 ***	−0.9030 ***
WML					−0.0223	−0.0245			0.0164	0.0172
F-value	66.94	66.63	55.63	55.17	50.62	48.22	31.17	29.32	26.19	24.46
R ²	0.180	0.182	0.248	0.253	0.248	0.253	0.262	0.269	0.262	0.269
R ² -adjusted	0.180	0.181	0.247	0.252	0.247	0.251	0.260	0.267	0.260	0.267

We use Fama and French’s (2015) five-factor models with a momentum factor to assess the differential exposure to various market factors between the ESG-tilted HSIESG and the parent index. The results allow an examination of the relative performance between the ESG-infused index and the parent index. The dependent variable is HSIESG return or HSI return. Independent variables are the market factors in the Fama–Macbeth equation. They are Mkt-RF market risk premium for developed international markets, SMB return differential between small- and large-cap stocks in developed international markets, SMB (size factor) return differential between small- and large-cap stocks in developed international markets, HML (P/B factor) return differential between high- and low-book-to-market stocks in developed international markets, RMW (profitability factor) return differential between stocks of high and low operating profitability in developed international markets, CMA (pro-growth factor) return differential between aggressive and conservative companies in developed international markets, and WML (momentum factor) winners minus losers in developed international markets. All time-series regressions use HAC standard errors of Newey and West (1984); a statistical significance of 1% is represented by ***. The sensitivity of the two index returns to systematic risk factors is similar.

Table 8. Sensitivity of HSIESG and HSI returns to systematic risk factors: Asia–Pacific markets excluding Japan.

	HSIESG	HSI	HSIESG	HSI	HSIESG	HSI	HSIESG	HSI	HSIESG	HSI
Full Sample										
Intercept	−0.0001	−0.0002	−0.0001	−0.0001	−0.0002	−0.0002	−0.0000	−0.0000	−0.0001	−0.0001
Mkt-RF	0.9793 ***	1.0079 ***	1.1317 ***	1.1519 ***	1.1063 ***	1.1269 ***	1.0595 ***	1.0774 ***	1.053 ***	1.0714 ***
SMB			0.1220	0.1301	0.0391	0.0485	−0.093	−0.0847	−0.1135	−0.1035
HML			0.6446	0.5894	0.7078	0.6516	0.4508 ***	0.4095 ***	0.4811 ***	0.4373 ***
RMW							−0.5372 ***	−0.5283 ***	−0.5177 ***	−0.5104 ***
CMA							−0.5325 ***	−0.5677 ***	−0.5221 ***	−0.5582 ***
WML					0.1591	0.1567			0.0530	0.0486
F-value	183.6	174.0	145.9	115.3	192.7	144.9	215.5	163.4	198.1	138.1
R ²	0.614	0.609	0.683	0.663	0.690	0.670	0.741	0.720	0.742	0.720
R ² -adjusted	0.613	0.608	0.682	0.662	0.689	0.669	0.740	0.719	0.741	0.719

We use Fama and French’s (2015) five-factor models with a momentum factor to assess the differential exposure to various market factors between the ESG-tilted HSIESG and the parent index. The results allow an examination of the relative performance between the ESG-infused index and the parent index. The dependent variable is HSIESG return or HSI return. Independent variables are the market factors in the Fama–Macbeth equation. They are Mkt-RF market risk premium for Asia–Pacific markets excluding Japan, SMB (size factor) return differential between small- and large-cap stocks in Asia–Pacific markets excluding Japan; HML (P/B factor) return differential between high- and low-book-to-market stocks in Asia–Pacific markets excluding Japan, RMW (profitability factor) return differential between stocks of high and low operating profitability in Asia–Pacific markets excluding Japan, CMA (pro-growth factor) return differential between aggressive and conservative companies in Asia–Pacific markets excluding Japan, and WML (momentum factor) winners minus losers in Asia–Pacific markets excluding Japan. All time-series regressions use the HAC standard errors of Newey and West (1984); a statistical significance of 1% is represented by ***. The results show that the change from the developed market to Asia–Pacific systematic risk factors produces almost the same findings.



Figure 1. Time-series plot of the daily observations of the levels of Hang Seng option-implied volatility index (VHSI) and the two stock indexes (HSIESG and HSI) for the period 8 September 2014–October 2021. The diagram shows the large variations in the perceived market volatility embedded in the Hang Seng Index options prices.

5. Conclusions

This study examines whether and to what extent the ESG-tilted weights change the performance of the index portfolio (HSIESG) relative to the parent index HSI. The paper shows that the daily returns of the two indexes are very highly correlated. However, the holding period return of the ESG-infused index is surprisingly higher than that of the parent index by over 67%. The unexpected result can be attributed to the difference in the strength of the asymmetrical volatility–return relationships between the two indexes for days with the highest and lowest volatility change, supporting our proposition that stocks with high ESG ratings are less susceptible to market volatility-induced trading pressures than firms with low ratings. The results also confirm our conjecture that ESG ratings are priced in the market, making high ESG companies less attractive to speculators, which increases these stocks’ tolerance against panic selling under market stress. Although our findings show that high-ESG stocks outperform their counterparts in market downturns, whether these companies can buffer against financial crises can be a theme for future studies by factoring in the implications of the volatility paradox discussed in Deghi et al. (2018).

These findings suggest that ESG information can play a role as a financial factor in the valuation process in financial management. Moreover, by improving a company’s ESG performance, the board and management may enhance the stock’s resilience to market volatility. The findings also have implications for the prospects of financial market development. Although our results are primarily based on the newly introduced HSIESG, it sheds light on the potential economic benefits of incorporating ESG information into the construction of stock market indexes. ESG indexes may support the development of relevant index products, such as ESG-linked equity ETFs, derivatives, other exchange-traded products, and mutual funds. Currently, there are 36 and 34 investment products linked to the parent indexes, i.e., the HSI and Hang Seng China Enterprises Index, respectively. These products include local ETFs, ETFs listed around the world, leveraged and inverse products in Hong Kong, classification funds and listed open-end funds in China, mandatory provident funds in Hong Kong, and index funds worldwide. Growth in such markets may help promote Hong Kong as a major sustainability and green financial hub and reinforce its status as a global financial center.

Author Contributions: J.K.W.F.: Conceptualization, Software, Methodology, Administration, Writing. F.Y.E.L.: Conceptualization, Data curation, Software, Methodology, Writing. Y.T.: Supervision, Writing, Methodology, Administration. All authors have read and agreed to the published version of the manuscript.

Funding: The Thematic Research Fellowship from the Hong Kong Institute for Monetary and Financial Research.

Data Availability Statement: The daily market factors for the Fama and French regression analysis are obtained from the Kenneth French library, and the rest of the dataset is downloaded from Bloomberg terminal.

Acknowledgments: We would like to thank K. C. Kwok, Giorgio Valente, Hongyi Chan, and seminar participants from the Hong Kong Monetary Authority for their comments and suggestions. Fung acknowledges support in the form of the Thematic Research Fellowship from the Hong Kong Institute for Monetary and Financial Research. The research provided by Lam was completed during his posting at the Hong Kong Institute for Monetary and Financial Research. This paper represents the views of the authors, which are not necessarily the views of the Hong Kong Monetary Authority, Hong Kong Institute for Monetary and Financial Research, or its board of directors or council of advisers. The aforementioned entities except the authors take no responsibility for any inaccuracies or omissions contained in the paper.

Conflicts of Interest: Author F. Y. Eric Lam was employed by the company Citigroup. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A

Table A1. The mean returns of the HSI ESG ($\Delta \ln \text{HSIESG}$) and HSI ($\Delta \ln \text{HSI}$) for each decile of volatility return ($\Delta \ln \text{VHSI}$) classification for the overall period and the prelaunch and postlaunch subperiods: A bootstrapping approach.

Decile	Overall			Prelaunch			Postlaunch		
	$\Delta \ln \text{VHSI}$	$\Delta \ln \text{HSIESG}$	$\Delta \ln \text{HSI}$	$\Delta \ln \text{VHSI}$	$\Delta \ln \text{HSIESG}$	$\Delta \ln \text{HSI}$	$\Delta \ln \text{VHSI}$	$\Delta \ln \text{HSIESG}$	$\Delta \ln \text{HSI}$
1	−9.0953	1.0496 ***	1.1261 ***	−8.5521	0.9139 ***	0.9788 ***	−10.0522	1.3110 ***	1.4060 ***
2	−4.8753	0.6018 ***	0.6380 ***	−4.7405	0.5751 ***	0.5965 ***	−5.1077	0.6366 ***	0.6991 ***
3	−3.3732	0.4504 ***	0.4737 ***	−3.3077	0.4153 ***	0.4319 ***	−3.4804	0.5012 ***	0.5462 ***
4	−2.3367	0.3418 ***	0.3534 ***	−2.2652	0.2854 ***	0.2977 ***	−2.4734	0.4751 ***	0.4931 ***
5	−1.2689	0.1447 ***	0.1519	−1.1902	0.2009 ***	0.2002 ***	−1.4111	0.0582	0.0628
6	−0.1699	0.0390	0.0515	−0.0936	0.0610	0.0663	−0.3035	0.0148	0.0408
7	1.0704	−0.0478	−0.0637	1.1353	−0.0483	−0.0587	0.9268	−0.0775	−0.1038
8	2.6172	−0.2573 ***	−0.2899 ***	2.6346	−0.1870	−0.2040	2.5615	−0.3746	−0.4274 ***
9	4.9207	−0.6085 ***	−0.6334 ***	4.7841	−0.5562 ***	−0.5867 ***	5.2033	−0.8251 ***	−0.8652 ***
10	12.7039	−1.6255 ***	−1.7235 ***	11.9064	−1.4779 ***	−1.5480 ***	14.0373	−1.7779 ***	−1.9015 ***

Daily index returns are grouped into 10 bins according to market volatility returns. The decile 1 bin contains the returns of the indexes on days with the greatest drop in market volatility, while the decile 10 bin contains the returns on days with the steepest rise in market volatility. The mean returns of the indexes are positive on days in the bins with a negative mean volatility return (i.e., deciles 1–6), while the mean returns of both indexes are positive on days in the bins with a positive mean volatility return. The results are consistent for the three periods. The 1% statistical significance level, indicated by ***, is determined via a bootstrapping method by resampling 10,000 times the returns. The positive mean returns for both indexes in the decile 10 bins are less than the magnitudes of the negative returns in the corresponding decile 1 bins, a result consistent for the overall period and the two subperiods. The findings show that an up jump in volatility has a greater impact on the index return than a down jump in volatility. Most importantly, the ESG-infused index has a significantly less asymmetrical volatility–return relationship than the parent index. For the overall period, the HSIESG has a lower mean return than the parent index (1.0496% vs. 1.1261%) for days with the highest drops in volatility; the opposite is true for days with the greatest spikes in volatility, and the HSIESG has a less negative mean return than the parent index (−1.6255% vs. −1.7235%).

Table A2. The differential mean returns between the HSI ESG and HSI ($\Delta \ln \text{HSI ESG} - \Delta \ln \text{HSI}$) for each decile of volatility return classification. A bootstrapping approach.

		Overall		Prelaunch		Postlaunch	
$\Delta \ln \text{VHSI}$		$\Delta \ln \text{HSI ESG} - \Delta \ln \text{HSI}$	$\Delta \ln \text{VHSI}$	$\Delta \ln \text{HSI ESG} - \Delta \ln \text{HSI}$	$\Delta \ln \text{VHSI}$	$\Delta \ln \text{HSI ESG} - \Delta \ln \text{HSI}$	
Decile							
1	−9.0953	−0.0765 ***	−8.5521	−0.0649 ***	−10.0522	−0.095 ***	
2	−4.8753	−0.0362 ***	−4.7405	−0.0214	−5.1077	−0.0625	
3	−3.3732	−0.0233	−3.3077	−0.0166	−3.4804	−0.045	
4	−2.3367	−0.0116	−2.2652	−0.0123	−2.4734	−0.018	
5	−1.2689	−0.0072	−1.1902	0.0007	−1.4111	−0.0046	
6	−0.1699	−0.0125	−0.0936	−0.0053	−0.3035	−0.026	
7	1.0704	0.0159	1.1353	0.0104	0.9268	0.0263	
8	2.6172	0.0326 ***	2.6346	0.017	2.5615	0.0528	
9	4.9207	0.0249 ***	4.7841	0.0305 ***	5.2033	0.0401	
10	12.7039	0.098 ***	11.9064	0.0701 ***	14.0373	0.1236 ***	

The 1% statistical significance level, indicated by ***, is determined via a bootstrapping method by resampling 10,000 times the returns. The results show that the HSI generally outperforms the HSI ESG for days with the largest volatility drop, as shown in the results from the decile 1 bin, while the opposite is true for days with the steepest rise in volatility, as shown in the results from the decile 10 bin. The above findings are similar for the two subperiods. As noted in Table 3, the less negative mean return of the ESG-infused index on days with the steepest rise in volatility produces a substantially higher holding period return than the parent index (i.e., 3.9657% vs. 2.369%) despite the seemingly minor and statistically insignificant difference in the standard deviation of daily returns (i.e., 0.0115% vs. 0.0119%). This result leads to the conclusion that the HSI ESG has a substantially higher holding period return than the parent index because the ESG-infused index has a significantly less negative mean return than the parent index on days with the highest volatility spikes (i.e., the decile 10 bin). Moreover, the positive mean return differential ($\Delta \ln \text{HSI ESG} - \Delta \ln \text{HSI} > 0$) in bin 10 is greater than the magnitude of the negative mean return differential ($\Delta \ln \text{HSI ESG} - \Delta \ln \text{HSI} < 0$) in bin 1. To understand the large difference in the overall holding period return between the two indexes, we calculate the cumulative returns for days included in volatility change in decile 1, decile 10, and the rest of the sample period. We find the following: (1) the cumulative return for days in decile 1 (the top 10% volatility change) is −380% and −399% for the HSI ESG and HSI, respectively; (2) the cumulative return for days in decile 10 (the bottom 10% volatility change) is −343% and −355% for the HSI ESG and HSI, respectively; (3) the cumulative returns for all other days (deciles 2–9, both deciles included) are 41.74% and 46.01% for the HSI ESG and HSI, respectively.

References

Authors, John. 2020. *ESG Investing Is Having a Good Crisis. It's Also Killing Jobs*, Bloomberg Markets. Paris: OECD.

Berg, Florian, Julian F. Kölbel, and Roberto Rigobon. 2022. Aggregate Confusion: The Divergence of ESG Ratings. *Review of Finance* 26: 1315–44. [CrossRef]

Bissoondoyal-Bheenick, Emawtee, Robert Brooks, and Hung Xuan Do. 2023. ESG and Firm Performance: The Role of Size and Media Channels. *Economic Modelling* 121: 106203. [CrossRef]

Brown Brothers Harriman (Hong Kong) Limited. 2019. *2019 Global ETF Investor Survey: Greater China Results Supplement*. Hong Kong: Brown Brothers Harriman.

Cao, Mingyao, Keyi Duan, and Haslindar Ibrahim. 2023. Local Government Debt and Its Impact on Corporate Underinvestment and ESG Performance: Empirical Evidence from China. *Sustainability* 15: 11116. [CrossRef]

Davoodi, Shayan, Ali Fereydooni, and Mohammad A. Rastegar. 2024. Can Portfolio Construction Considering ESG Still Gain High Profits? *Research in International Business and Finance* 67: 102126. [CrossRef]

Deghi, Andrea, Peter Welz, and Dawid Zochowski. 2018. A new financial stability risk index to predict the near-term risk of recession. *Financial Stability Review*. Available online: https://www.ecb.europa.eu/pub/financial-stability/fsr/special/html/ecb.fsrart201805_1.en.html (accessed on 14 December 2023).

Dunn, Jeff, Shaun Fitzgibbons, and Lukasz Pomorski. 2017. Assessing Risk through Environmental, Social and Governance Exposures. *Journal of Investment Management* 16: 4–17.

Eccles, Robert G., Ioannis Ioannou, and George Serafeim. 2014. The Impact of Corporate Sustainability on Organizational Processes and Performance. *Management Science* 60: 2835–57. [CrossRef]

El Ghouli, Sadok, Omrane Guedhami, Chuck C. Y. Kwok, and Dev R. Mishra. 2011. Does Corporate Social Responsibility Affect the Cost of Capital? *Journal of Banking and Finance* 35: 2388–406. [CrossRef]

Fama, Eugene F., and Kenneth R. French. 2015. A Five-Factor Asset Pricing Model. *Journal of Financial Economics* 116: 1–22. [CrossRef]

Fatemi, Ali, Iraj Fooladi, and Hassan Tehrani. 2015. Valuation Effects of Corporate Social Responsibility. *Journal of Banking & Finance* 10: 182–92.

- Friede, Gunnar, Timo Busch, and Alexander Bassen. 2015. ESG and Financial Performance: Aggregated Evidence from More than 2000 Empirical Studies. *Journal of Sustainable Finance & Investment* 5: 210–33.
- Giese, Guido, Linda E. Lee, Dimitris Melas, Zoltán Nagy, and Laura Nishikawa. 2019a. Foundations of ESG Investing: How ESG Affects Equity Valuation, Risk, and Performance. *Journal of Portfolio Management* 45: 1–15. [CrossRef]
- Giese, Guido, Linda E. Lee, Dimitris Melas, Zoltán Nagy, and Laura Nishikawa. 2019b. Performance and Risk Analysis of Index-Based ESG Portfolios. *Journal of Index Investing* 9: 46–57. [CrossRef]
- Godfrey, Paul C., Craig B. Merrill, and Jared M Hansen. 2009. The Relationship between Corporate Social Responsibility and Shareholder Value: An Empirical Test of the Risk Management Hypothesis. *Strategic Management Journal* 30: 425–45. [CrossRef]
- Gregory, Alan, Rajesh Tharyan, and Julie Whittaker. 2014. Corporate Social Responsibility and Firm Value: Disaggregating the Effects on Cash Flow, Risk and Growth. *Journal of Business Ethics* 124: 633–57. [CrossRef]
- Hong, Harrison, and Marcin Kacperczyk. 2009. The Price of Sin: The Effects of Social Norms on Markets. *Journal of Financial Economics* 93: 15–36. [CrossRef]
- International Monetary Fund. 2019. *Global Financial Stability Report: Lower for Longer*. Washington, DC: International Monetary Fund (IMF).
- Jin, Yuqian, Qingfu Liu, Yuman Tse, and Kaixin Zheng. 2023. Hedging COVID-19 Risk with ESG Disclosure. *International Review of Economics and Finance* 88: 27–46. [CrossRef]
- Krueger, Philipp. 2015. Corporate Goodness and Shareholder Wealth. *Journal of Financial Economics* 115: 304–29. [CrossRef]
- Li, Kunming, Linxing Huang, Jinshan Zhang, Zhencheng Huang, and Liting Fang. 2023. Can ESG Performance Alleviate the Constraints of Green Financing for Chinese Enterprises: Evidence from China's A-Share Manufacturing Companies. *Sustainability* 15: 10970. [CrossRef]
- Lins, Karl V., Henri Servaes, and Ane Tamayo. 2017. Social Capital, Trust, and Firm Performance: The Value of Corporate Social Responsibility during the Financial Crisis. *Journal of Finance* 72: 1785–824. [CrossRef]
- Melas, Dimitris, Zoltán Nagy, and Padmakar Kulkarni. 2018. *Factor Investing and ESG Integration, Factor Investing: From Traditional to Alternative Risk Premia*. London: ISTE Press; Oxford: Elsevier, pp. 389–413.
- Moody's. 2020. *Beyond Passive, ESG Investing Is the Next Growth Frontier for Asset Managers, Moody's Investor Service*. New York: Moody's.
- Nagy, Zoltán, Altaf Kassam, and Linda E. Lee. 2016. Can ESG Add Alpha? An Analysis of ESG Tilt and Momentum Strategies. *Journal of Investing* 25: 113–24.
- Newey, Whitney K., and Kenneth D. West. 1984. A Simple, Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55: 703–8. [CrossRef]
- Oikonomou, Ioannis, Chris Brooks, and Stephen Pavelin. 2012. The Impact of Corporate Social Performance on Financial Risk and Utility: A Longitudinal Analysis. *Financial Management* 41: 483–515. [CrossRef]
- Pulino, Silvia C., Mirella Ciaburri, Barbara S. Magnanelli, and Luigi Nasta. 2022. Does ESG Disclosure Influence Firm Performance? *Sustainability* 14: 7595. [CrossRef]
- Rojo-Suárez, Javier, and Ana B. Alonso-Conde. 2024. Have Shifts in Investor Tastes Led the Market Portfolio to Capture ESG Preferences? *International Review of Financial Analysis* 91: 103019. [CrossRef]
- Ward, Karen, and Jennifer Wu. 2019. *How Is ESG Affecting the Investment Landscape? (On the Minds of Investors)*. New York: J.P. Morgan Asset Management.
- Wasiuzzaman, Shaista, Salihu A. Ibrahim, and Farahiyah Kawi. 2022. Environmental, Social and Governance (ESG) Disclosure and Firm Performance: Does National Culture Matter? *Meditari Accountancy Research* 31: 1239–65. [CrossRef]
- Wu, Jennifer, and Vincent Juvyns. 2020. *COVID-19 Shows ESG Matters More Than Ever (On the Minds of Investors)*. New York: J.P. Morgan Asset Management.
- Zhang, Ning, Yue Zhang, and Zhe Zong. 2023. Fund ESG Performance and Downside Risk: Evidence from China. *International Review of Financial Analysis* 86: 102526. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

A Survey of Literature on the Interlinkage between Petroleum Prices and Equity Markets

Miramir Bagirov ¹ and Cesario Mateus ^{2,*}

¹ School of Accounting, Finance and Economics, University of Greenwich, London SE10 9LS, UK; m.bagirov@greenwich.ac.uk

² Business School, Aalborg University, Fibigerstræde 11-73, 9220 Aalborg, Denmark

* Correspondence: cmateus@business.aau.dk

Abstract: The multifaceted interrelationship between petroleum prices and equity markets has been a subject of immense interest. The current paper offers an extensive review of a plethora of empirical studies in this strand of literature. By scrutinising over 190 papers published from 1983 to 2023, our survey reveals various research themes and points to diverse findings that are sector- and country-specific and contingent on employed methodologies, data frequencies, and time horizons. More precisely, petroleum price changes and shocks exert direct or indirect effects dictated by the level of petroleum dependency across sectors and the country's position as a net petroleum exporter or importer. The interlinkages tend to display a time-varying nature and sensitivity to major market events. In addition, volatility is not solely spilled from petroleum to equity markets; it is also observed to transmit in the reverse direction. The importance of incorporating asymmetries is documented. Lastly, the summarised findings can serve as the basis for further research and reveal valuable insights to market participants.

Keywords: petroleum prices; aggregate equity market returns; equity sector returns; petroleum price shocks; volatility transmission; petroleum exporters; petroleum importers

Citation: Bagirov, Miramir, and Cesario Mateus. 2024. A Survey of Literature on the Interlinkage between Petroleum Prices and Equity Markets. *Journal of Risk and Financial Management* 17: 40. <https://doi.org/10.3390/jrfm17010040>

Academic Editor: Thanasis Stengos

Received: 14 December 2023

Revised: 17 January 2024

Accepted: 18 January 2024

Published: 22 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Petroleum is widely regarded as one of the critical components driving the economies of nations that export and import this strategic fossil fuel (Gupta 2008; Korhonen and Ledyeva 2010; Demirer et al. 2015). Notwithstanding the decline in global primary energy consumption in 2021, petroleum continues to remain the major source of energy, fulfilling more than 30 percent of demand.¹ The prices of petroleum have traditionally been more volatile in comparison to those of other commodities. Any factors that cause considerable petroleum demand and supply imbalances subsequently lead to shocks in the petroleum market (Hamilton 2003; Kilian 2009; Kilian and Park 2009). The recent financialization of commodity markets resulted in a strengthened degree of interrelationship between petroleum and stock markets (Tang and Xiong 2012; Silvennoinen and Thorp 2013). Several incidents of unprecedented petroleum price fluctuations over the past two decades overlapped with analogous shifts in stock markets, thereby pointing to probable financial contagion. It is therefore unsurprising that information arising from the petroleum market is of utmost relevance to market participants.

The literature has documented different theoretical interaction mechanisms between petroleum prices and stock markets. Degiannakis et al. (2018), in their study, discuss five transmission channels, namely stock valuation, monetary, output, fiscal, and uncertainty. The present work's interest lies in the first, as it is widely regarded as a direct and crucial channel that contributes to the elucidation of the rationale for the linkage between the prices of petroleum and stocks. The valuation approaches suggest that stock prices are equal to the present value of expected future cash flows discounted by employing a required rate

of return (Huang et al. 1996). Hence, factors that lead to alterations in discount rates and cash flows, as a result, affect stock prices. The cash flows of companies can be influenced positively or negatively by the price of petroleum, which is contingent in their positions as petroleum consumers or producers (Mohanty and Nandha 2011; Basher et al. 2012). Specifically, the petroleum price hikes elevate the costs of producing goods, given the unavailability of full substitution effects, and lessen expected cash flows for companies consuming petroleum (Basher and Sadorsky 2006; Smyth and Narayan 2018). However, they generally increase future cash flows for companies producing petroleum (Basher et al. 2018; Degiannakis et al. 2018). Furthermore, expected discount rates are also impacted by the price of petroleum as they tend to mirror the state of the economy (Mohanty and Nandha 2011), which is in turn sensitive to shocks in the petroleum market (Cunado and Perez de Gracia 2005; Korhonen and Ledyeva 2010; Cashin et al. 2014; among others). Thus, swings in the price of petroleum may exercise direct and indirect effects on stock prices through their effects on cash flows and discount rates, respectively.

Ever since the studies of Hamilton (1983), Jones and Kaul (1996), and Huang et al. (1996), an extensive body of literature has emerged devoted to the investigation of the impacts of petroleum price swings on macroeconomic variables (see, among others, Burbidge and Harrison 1984; Mork 1989; Hooker 1996; Bjornland 2000; Cunado and Perez de Gracia 2003; Lardic and Mignon 2006; Korhonen and Ledyeva 2010; Hou et al. 2016; Lorusso and Pieroni 2018) and equity markets (see, among others, Sadorsky 2001; Papapetrou 2001; Basher and Sadorsky 2006; Nandha and Faff 2008; Miller and Ratti 2009; Narayan and Narayan 2010; Lee et al. 2012; Basher et al. 2012; Silvapulle et al. 2017). Others direct their attention towards examining the interactions of volatilities and shocks between petroleum prices and stock markets, focusing on developed markets (Apergis and Miller 2009; Vo 2011; Mensi et al. 2013; Chang et al. 2013; Guntner 2014; Khalfaoui et al. 2015; Salisu and Oloko 2015; Angelidis et al. 2015; Ewing and Malik 2016; among others) and emerging markets (Malik and Hammoudeh 2007; Arouri et al. 2011a; Fang and You 2014; Lin et al. 2014; Koh 2017; Yousaf and Hassan 2019; among others). Furthermore, a subset of the preceding group distinguishes between petroleum exporters and importers (Wang et al. 2013; Wang and Liu 2016; Basher et al. 2018; Ashfaq et al. 2019; Sarwar et al. 2019; Belhassine and Karamti 2021; Enwereuzoh et al. 2021; among others) and accentuates the significance of the sectoral analysis (Kilian and Park 2009; Malik and Ewing 2009; Elyasiani et al. 2011; Arouri et al. 2012; Degiannakis et al. 2014; Broadstock and Filis 2014; Kang et al. 2016; Bouri et al. 2016; Belhassine 2020; Umar et al. 2020; Hwang and Kim 2021; Mishra and Mishra 2021; among others). More recent studies investigate the impact of the COVID-19 outbreak on the nexus between petroleum prices and equity markets (Hung and Vo 2021; Bouri et al. 2021; Zhang et al. 2021; Zhang and Hamori 2021; Benlagha and El Omari 2022; Jebabli et al. 2022; among others).

In light of the foregoing, the present survey strives to synthesise and consolidate an extensive volume of empirical studies that constantly grow, thereby facilitating a holistic comprehension of the interrelation between petroleum prices and stock markets and the identification of new directions for extending research in this strand of literature. To this end, the procedure of obtaining the final list of studies was commenced by identifying leading journals that systematically make publications in the area of interest. The rankings compiled by the Australian Business Deans Council are used for journal quality filtering. All relevant published papers were searched within the selected journals, which are from the A*, A, and B rating categories, without constraining ourselves to any specific period. This approach generated a vast volume of papers. In order to make our list concise and manageable, we focused our attention on the most influential works. It is worth mentioning that some studies with important contributions published in lower-ranked journals were also considered. The final sample includes 194 papers published from 1983 to 2023,² out of which 83% have A* and A ratings, while the remaining 12% and 5% have B and lower ratings, respectively, which allows one to observe how research interest has evolved over time. The main differentiation of the current paper from others is that, by reviewing nine

research themes, it splits findings across multiple country groups with various levels of market development, thereby uncovering valuable insights that can help academics and market participants develop a deeper comprehension of the petroleum-equity market nexus and pointing out a few directions for expanding the existing research.

In the review, each section is devoted to one prominent theme, where papers are categorised by country groups within paragraphs. This framework provides a complete picture, given that sensitivities to petroleum prices tend to be dissimilar across developed, emerging, and frontier markets. We commence the next section by discussing the effects of petroleum prices on various macroeconomic indicators. Section 3 analyses the impact of changes in petroleum prices on equity returns at the aggregate level, while Section 4 assesses the relationship at the sector level. Section 5 focuses on the role of forces driving petroleum prices in explaining the behaviour of equity returns. Section 6 assesses the time-varying connections between petroleum prices and equity markets. Volatility interactions between petroleum prices and equity markets at the aggregate and sector levels are inspected in Section 7. Section 8 emphasises the importance of distinguishing between petroleum-exporting and importing countries. Section 9 concentrates on the impact of petroleum prices on firm-level equity returns. Section 10 evaluates intraday linkages between petroleum prices and equity markets. The final section concludes the survey and outlines prospective avenues for future research.

2. Petroleum Prices and Economic Variables

The debates in relation to the effects of petroleum price shocks on macroeconomic variables began in the 1970s of the last century.³ Hamilton's (1983) empirical study, which was one of the earliest works to investigate such interconnections, induced discussions in this area. The author investigated the impact of changes in the price of petroleum on the economy of the United States for the period 1948–1980. Hamilton (1983) found that shocks associated with petroleum prices were a promoting factor in the majority of the post-war United States recessions. Burbidge and Harrison (1984) studied the effects of increases in the petroleum price on the domestic economies of Canada, Germany, Japan, the UK, and the US. The authors used monthly data from 1961 to 1982 and applied the vector autoregression (VAR) model. Their empirical findings suggest that the economic variables of these countries were influenced by petroleum price innovations. However, the effects of the 1973–1974 petroleum price shocks were stronger compared to the 1979–1980 set of shocks, where the effects were minimal in all cases apart from Japan. Gisser and Goodwin (1986) examined the influence of petroleum prices on the US economy for the period 1961–1982. The authors found that macroeconomic indicators were significantly affected by crude petroleum prices. Hooker (1996) analysed the relationship between petroleum prices and US macroeconomic indicators. The author concluded that the effect of the 1973 petroleum price shock on the macro-economy was substantial and measured well. In addition, Hooker (1996) highlighted the significance, but at the same time the insufficiency, of the 1979 petroleum price shock to reflect the 1980–1982 recession's dynamics. The author's investigation of the late 1980s suggests that the simple increases and decreases in price do not fully represent the relationship between the price of petroleum and the macro-economy. Hamilton's (1996) study indicates that the impact of petroleum price increases on GDP was significant from 1948 to 1973, in contrast to the later period from 1973 to 1994. The empirical work of Bjørnland (2000) analysed the impacts of petroleum price shocks on the macroeconomic indicators of four European countries, namely Germany, Norway, the United Kingdom, and the United States. The findings indicate that in all cases, except Norway, the effects of petroleum price shocks are significantly negative.

A number of empirical studies, through the application of asymmetric or nonlinear approaches, examined the relationship between petroleum prices and macroeconomic variables. Mork (1989) extended Hamilton's study conducted in 1983 in order to include the petroleum market collapse in 1985–1986. The author particularly paid attention to the probability of asymmetric reactions to both rises and declines in the price of petroleum.

Mork (1989) confirmed that the link between increases in the price of petroleum and economic activity is negative. On the other hand, the findings suggest that the effects of declines in the price of petroleum have no significant impacts but are different from petroleum price increases. Mork et al. (1994) studied the reactions of macroeconomic variables in seven countries, namely Canada, France, Germany, Japan, Norway, the United Kingdom, and the United States, to increases and decreases in the price of petroleum. Their empirical findings indicate a strong asymmetry in the reactions. The responses of the majority of countries to increases in the price of petroleum are significant and negative. The reactions to decreases in the price of petroleum are mainly positive, but significant in the case of Canada and the United States. The findings related to Norway differ in the sense that the effects of petroleum price rises and declines are positive and negative, respectively. Hamilton (2003) investigated the nonlinear relationship between changes in the price of petroleum and the gross domestic product (GDP) of the United States for the period 1949–2001. The author's empirical findings clearly indicate signs of nonlinearity. Hamilton (2003) reports that increases in the price of petroleum have more importance compared to decreases in the price of petroleum.

The more recent empirical works on the macroeconomic effects of petroleum price fluctuations have obtained various outcomes. Some studies consider developed economies. Cunado and Perez de Gracia (2003) analysed the relationship between petroleum prices and macroeconomic variables, such as inflation and industrial production indexes, in fifteen European countries. The authors employed quarterly data covering the period from 1960 to 1999 and used various proxies of petroleum price shocks. Their main findings suggest that the prices of petroleum have a constant impact on inflation. However, the effects on production growth rates are short-term and asymmetric. In addition, the influence is greater when petroleum prices are measured in local currencies. The results also indicate that the reactions of fifteen European countries to shocks significantly differ. Jimenez-Rodriguez and Sanchez (2005) assessed the impacts of shocks arising from petroleum prices on the real economic activity of Canada, France, Germany, Italy, Japan, Norway, the UK, the US, and the Euro area. The authors employed both linear and non-linear models to carry out the multivariate VAR analysis. The authors found that the majority of economies are more negatively influenced by petroleum price increases than by declines, which aligns with previous findings on asymmetric effects. Furthermore, increases in the price of petroleum negatively influence the economic activity of all petroleum-importing countries, apart from Japan. However, reactions between petroleum-exporting countries vary. The impact of petroleum price shocks on the UK is negative, while in the case of Norway, the results demonstrate a positive effect. Lardic and Mignon (2006), using quarterly data from 1970 to 2003, studied the long-term interrelationship between petroleum prices and the GDP of twelve European countries. The authors based their investigation on asymmetric cointegration. They found the existence of asymmetric cointegration between petroleum prices and the GDP in most cases. In addition, the results reject standard cointegration. Lardic and Mignon (2008) extended their work to the economies of the G7, the United States, Europe, and the Euro area. Their empirical findings confirm the presence of asymmetric cointegration between petroleum prices and GDP in all studied countries. Hou et al. (2016) studied the impact of petroleum price shocks on the macroeconomic variables of a petroleum exporting economy, Canada, for the period 1980–2011. Their findings show that the effect of petroleum shocks on the aggregate demand of Canada is stimulative. In addition, shocks appreciate the Canadian national currency and lead to an improvement in terms of trade and a reduction in real wages. Lorusso and Pieroni (2018) assess the influence of fluctuations in the price of petroleum on the economy of the United Kingdom for the period from 1976 to 2014. The outcomes show that the repercussions of petroleum price movements on the macroeconomic variables of the United Kingdom are contingent on various types of petroleum shocks. Wen et al. (2021) concentrate on investigating the responses of inflation in the G7 countries to petroleum demand, supply, and risk shocks over the period from 1997 to 2019. The authors observe that petroleum

shocks exert diverse effects on inflation in the studied countries, with the US displaying the largest sensitivity.

Other works utilise a sample of developed and emerging economies. Cunado and Perez de Gracia (2005) investigate the effects of petroleum price shocks on consumer price indexes and the economic activity of Asian countries, namely Japan, Malaysia, the Philippines, Singapore, South Korea, and Thailand. The authors used quarterly data for the period from 1975 to 2002. The results show that both consumer price indexes and economic activity are significantly affected by petroleum price shocks, although the effects are short-term and greater when petroleum prices are represented in local currencies. Furthermore, the authors found the existence of asymmetries in the relationship between petroleum prices and macroeconomic variables for some of the countries. Kim et al. (2017) analyse the effect of positive shocks arising from petroleum prices on the economy of China, where the key focus is on the interest rate reaction to shocks. The authors used monthly data from January 1992 to May 2014 and applied structural vector autoregression (SVAR), time-varying parameter structural vector autoregression (TVP SVAR), and generalised impulse response vector autoregression (GIR VAR) econometric models. They found that the reaction of the country's interest rate to petroleum price shocks varies over time and demonstrates different signs. Specifically, the interest rate response to the shocks between April 1992 and October 2001 is negative, while the reaction between November 2001 and May 2014 is positive. Liu et al. (2020) provide evidence of negative responses of the money supply and economic growth in China to petroleum price shocks, whereas the impact on inflation is found to be positive. Lin et al. (2023) show that the effects of petroleum price shocks on macroeconomic variables in Euro-19, China, Japan, and the US are time-varying and heterogeneous, which amplify over the turbulent period.

A group of studies focuses on petroleum exporters and importers. Korhonen and Ledyeva (2010) investigate the impact of shocks emerging from petroleum prices on the economies of both petroleum-producing and petroleum-consuming countries. The authors distinguished two types of effects, direct and indirect, of petroleum price shocks in their study. The key results suggest that the direct effects of increases in the price of petroleum are positive for petroleum-exporting countries, but the indirect effects are negative, although minor. The majority of petroleum-importing countries are adversely affected by shocks in the price of petroleum. However, some economies demonstrated persistence in responding to petroleum price changes. Nusair (2016) examines the impact of petroleum price shocks on the economies of the Gulf Cooperation Council (GCC), namely Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, and the UAE. The author applied the nonlinear cointegrating autoregressive distributed lag model to analyse the short-term and long-term effects. The findings demonstrate the existence of asymmetries in all instances. Precisely, positive movements in the price of petroleum contribute to the growth in real GDP. Conversely, the results show that negative changes in the price of petroleum are only significant for two countries, Kuwait and Qatar, implying that declines depress real GDP. Overall, the empirical outcomes suggest that the impact of positive petroleum price fluctuations on real GDP is considerably greater in contrast to negative movements. Lee et al. (2017) study the interconnection between petroleum price shocks and country risks in petroleum-importing and exporting countries. Their empirical findings indicate that petroleum price shocks significantly influence country risks in petroleum exporting and importing economies, but the effects vary. Smiech et al. (2021) document that the industrial production in petroleum exporting countries, namely Canada, Mexico, Norway, and Russia, is declined by the petroleum price uncertainty shocks. Furthermore, the authors report immediate currency depreciations that are long-lasting for Mexico and Russia. Tan and Uprasen (2023) observe that in petroleum-exporting countries, higher petroleum prices lead to a decline in income inequality, but for petroleum-importing countries, they contribute to an increase in income inequality. Conversely, upsurges in petroleum price volatility are found to increase income inequality in both groups.

3. Petroleum Prices and Aggregate Stock Markets

A large volume of empirical work explores the impact of petroleum prices on equity returns at the aggregate level.⁴ Some studies have found a negative interrelationship between petroleum prices and stock markets. Jones and Kaul (1996), who made the initial contribution to this field, analyse the effects of changes in petroleum prices on several international equity markets. The authors, through the application of the standard cash flow dividend valuation model, found that the effects of petroleum price changes on current and future corporate cash flows could elucidate the changes in equity returns. Specifically, their findings show that equity market returns are negatively affected by petroleum shocks. Sadorsky (1999) investigates interactions between changes in the price of petroleum and equity returns in the United States. The author applied the vector autoregression model and used monthly data for the period 1947–1996. Sadorsky (1999) found that equity returns are significantly depressed by petroleum price increases. Furthermore, the author's results indicate that the effects of petroleum on equity returns were stronger during the period from 1986 to 1996, when petroleum prices experienced large declines. Papapetrou (2001) studies the interconnections between petroleum prices and stock returns in Greece. The author employed monthly data from 1989 to 1999 and applied the multivariate vector autoregressive model. Papapetrou (2001) obtained similar findings, where positive petroleum price shocks negatively affect stock returns. Driesprong et al. (2008) examine the relationship between movements in petroleum prices and stock returns in 48 developed and emerging countries across the world. Their findings indicate that future stock returns are drastically lowered by increases in the price of petroleum. Chen (2010), by the application of time-varying transition probability Markov-switching models analyse the relationship between petroleum prices and the US stock market, represented by the S&P500 index, for the period from 1957 to 2009. The author focuses on whether an increase in the price of petroleum causes recessions in the US stock market. The empirical findings provide strong evidence that the equity market is pushed into bear territory by increasing petroleum prices. Chen (2010) also found that higher petroleum prices increase the probability of staying in a bear regime, but evidence for such outcomes is weaker. Filis (2010) studies the interconnections between prices of petroleum, which is represented by the Brent grade, industrial production, the consumer price index, and the stock market in Greece. The author applied the vector error correction model (VECM) and VAR models and used data covering the period from 1996 to 2008. Filis (2010) found that the influence of petroleum prices on the stock market in Greece is significantly negative. Basher et al. (2012) examine the dynamic relationship between West Texas Intermediate petroleum prices, emerging market equity prices, represented by the Morgan Stanley Capital International (MSCI) emerging markets index, and exchange rates. The authors employed monthly data for the period of 1988–2008 and estimated the structural vector autoregression model. Their empirical findings indicate that in the short run, exchange rates and emerging market equity prices are depressed by positive shocks to the prices of petroleum. Asteriou and Bashmakova (2013) investigate the interrelationship between the petroleum price risk and equity market returns of ten emerging countries from the central and eastern European regions, namely the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Russia, Slovakia, and Slovenia. The authors employed daily data covering the period from 1999 to 2007 and applied the international multi-factor model. Their findings indicate that equity market returns are sensitive to petroleum price risk. In particular, the petroleum price risk is statistically significant and negative, implying that the price of petroleum is an important element in determining equity returns, and increases in the price level depress equity market returns. A primary explanation of the negative effects caused by petroleum price rises is that the studies have predominantly considered countries with a relatively high level of petroleum dependency.

Other empirical works have identified a positive relationship between the prices of petroleum and stock markets. Narayan and Narayan (2010), through the employment of daily data, investigate the relationship between petroleum prices and equity prices in

Vietnam for the period from 2000 to 2008. The authors found that the impact of petroleum prices on equity prices is positive and statistically significant, which is not in line with the expected theoretical outcome. They suggest that unique factors, among which are an increase in foreign investment inflows and a change in investment preferences towards stocks that occurred during swift upsurges in the price of petroleum, had the dominant impact on the Vietnamese stock market and elucidate the presence of this interlinkage. Zhu et al. (2011) examine the relationship between petroleum prices and stock markets in fourteen OECD and non-OECD countries using monthly data from 1995 to 2009. Their empirical findings indicate the existence of a long-run bidirectional relationship between these variables. In addition, the results show that increases in crude petroleum prices positively affect stock prices. The authors explain the observed positive interrelation by citing the prevailing impact of industrial production and increased leveraged investment in equities as opposed to crude petroleum price rises. Zhu et al. (2014) analyse the dynamic dependence between prices of crude petroleum and equity markets in countries from the Asia-Pacific region, namely Australia, China, Hong Kong, India, Indonesia, Japan, South Korea, Malaysia, Singapore, and Taiwan. The authors use daily data covering the period from 2000 to 2012 and apply both conditional and unconditional copula models. Their empirical outcomes demonstrate that prior to the financial crisis, the dependence between petroleum prices and equity returns in most cases was weak and positive, which could potentially be attributed to the growth of economies in the Asia-Pacific region. However, in the post-crisis period, it strengthened substantially. Silvapulle et al. (2017), using monthly data for the period from 1999 to 2015, studied the relationship between petroleum prices and stock market indices of major petroleum-importing countries, such as China, France, Germany, India, Italy, Japan, Singapore, South Korea, Spain, and the United States. They found that the impact of petroleum prices on stock market indices is significant and positive, with the coefficient being higher during the pre-financial crisis period compared to the post-financial crisis period. The authors provide several elucidations for the detected findings, namely speculative behaviour in the petroleum market, the more dominant impact of alternative energy sources that the studied countries are dependent on, leveraged investment in equities, and the positive and cyclical association of equity market indices and petroleum prices with the global need for industrial resources.

A group of empirical studies have obtained mixed results or found no interrelationships between petroleum prices and stock markets. Huang et al. (1996) investigate the relationship between returns on petroleum futures and US equities. The authors used daily data and applied the VAR model. Their findings suggest that the impact of petroleum futures returns on US stock market returns, represented by the S&P 500 index, is non-existent. On the other hand, the authors found that some individual petroleum equity returns are led by petroleum futures returns. Basher and Sadorsky (2006), through the application of the international multi-factor model, examine the effects of fluctuations in the price of petroleum on stock market returns in emerging economies, namely Argentina, Brazil, Chile, Colombia, India, Indonesia, Israel, Jordan, Korea, Malaysia, Mexico, Pakistan, Peru, Philippines, Poland, South Africa, Sri Lanka, Taiwan, Thailand, Turkey, and Venezuela, for the period 1992–2005. The authors found strong evidence indicating that the petroleum price risk affects equity market returns in emerging economies, although the exact interrelation depends on the frequency of employed data. Cong et al. (2008) investigate the effects of changes in the prices of petroleum, represented by the Brent grade, on equity market returns in China for the period from 1996 to 2007. Their empirical results show no evidence of the statistically significant effects on returns of the majority of stock market indices in China, apart from the manufacturing index and some individual petroleum companies. O'Neill et al. (2008) studied interconnections between fluctuations in petroleum prices and equity market indices of Australia, Canada, France, the United Kingdom, and the United States. The authors found that the effects of high petroleum prices vary among the studied countries. The equity market returns of France, the United Kingdom, and the United States are adversely affected by higher petroleum prices, while the impacts are positive for

Australia and Canada. Miller and Ratti (2009) apply the vector error correction model to study the relationship between the price of crude petroleum and the equity markets of six countries, namely France, Germany, Italy, the United Kingdom, Canada, and the United States, over the period 1971–2008. The authors found clear evidence of the interrelationship for the period from 1971 to 1980 and 1988 to 1999 between the price of crude petroleum and the equity markets of all countries. On the other hand, their results show the absence of a statistically significant relationship between these variables during the period from 1980 to 1988. Furthermore, the findings indicate the disintegration of the relationship after 1999. Jammazi and Aloui (2010), through the combination of wavelet analysis and the Markov switching VAR (MS-VAR) model, assess the relationship between changes in petroleum prices and stock market returns in France, Japan, and the UK, for the period 1989–2007. The authors found that the response of the equity market variables to movements in the price of crude petroleum is negative and temporary during the moderate, for France, and expansion, for France and the UK, stock market phases. However, no effects are observed, except for Japan, during the recessionary stock market phase. In addition, their findings show that the negative interrelation was more pronounced prior to 1999. The empirical study of Ajmi et al. (2014) is focused on the investigation of the nonlinear relationships between petroleum prices, which are represented by both the WTI and Brent grades, and the stock markets of eleven countries from the Middle East and North Africa region over the period 2007–2012. The authors found that the interaction between petroleum prices and considered stock markets is nonlinear. In addition, the results suggest that the nonlinear causality is more evident in the case of the Brent grade. Narayan and Gupta (2015) used an extensive range of monthly historical data to study the relationship between petroleum prices and equity returns in the US, represented by the WTI grade and the S&P 500 index, respectively. The authors found evidence of asymmetric effects, that is, negative movements in petroleum prices in comparison to positive changes, are more important in predicting US equity returns. Hatemi et al. (2017) investigate the impact of petroleum prices on equity prices in the G7 countries and the world market for the period from 1975 to 2013. Their empirical findings obtained based on the symmetric causality test provide no evidence of the effects on equity prices in both G7 and world markets. However, when the authors apply the asymmetric test, the results indicate that increases in the price of petroleum induce rises in equity prices in Japan, the US, and the world market, whereas decreases in the price of petroleum depress equity prices in Germany. The reported mixed findings point to the fact that stock markets in the studied countries are heterogeneous in their reactions, given dissimilar levels of dependence on petroleum, and the importance of distinguishing the type of petroleum price movements.

4. Petroleum Prices and Stock Sectors

The aforementioned empirical works investigate the petroleum-stock market relationship using aggregate market indices. Some of the major limitations of aggregation are the following: (i) the weight of each industry in a market index varies, and hence, it tends to be biased towards certain industries (Mateus et al. 2017); (ii) sectors are not identical, and various factors can have dissimilar impacts (Faff and Brailsford 1999); and (iii) movements in the price of petroleum differently influence sectors (Nandha and Brooks 2009), given varying degrees of reliance on petroleum. Thus, one can naturally anticipate that reactions, in terms of magnitude and sign, can be nonuniform.⁵ A vast number of studies have conducted sectoral investigations in the context of developed markets. Faff and Brailsford (1999), employing the augmented market model, study the sensitivity of industry stock returns in Australia to petroleum prices for the period from 1983 to 1996. The authors use, in their analysis, monthly data on twenty-four Australian industries. Faff and Brailsford (1999) observe a significant positive sensitivity for industries such as Diversified Resources and Oil and Gas. In the case of the Transport and Paper and Packaging industries, they found significant negative responsiveness. Sadorsky (2001) examines the interrelationship between different risk factors, including crude petroleum prices, and

equity prices in the Oil and Gas industry in Canada. The author applied the multifactor market model and employed monthly data that covered the period from 1983 to 1999. The empirical findings indicate that petroleum prices influence the equity returns in the studied industry. Particularly, the hikes in petroleum prices increase equity returns of Canadian Oil and Gas companies. Boyer and Filion (2007) also found that rises in petroleum prices have a positive influence on the equity returns of Oil and Gas firms in Canada. Hammoudeh and Li (2005) investigate the sensitivity of US industry equity indices to petroleum, namely the Transportation and Petroleum industries, represented by the NYSE Transportation Index and Amex Petroleum Index, respectively. The authors found that in the case of the Petroleum industry, the sensitivity is positive but negative for the Transportation industry, with the first industry demonstrating the highest responsiveness. El-Sharif et al. (2005) analyse the relationship between the price of crude petroleum and stock values in the Oil and Gas industry of the United Kingdom. The authors found that Oil and Gas stock returns are affected by changes in crude petroleum prices. In particular, petroleum price upsurges tend to enlarge equity returns in the considered sector of the United Kingdom. Lee et al. (2012) investigate the interrelationship between movements in petroleum prices and sector equity indices of the G7 countries. The authors employed monthly data from 1991 to 2009 and applied the unrestricted VAR model in their study. Their empirical outcomes indicate that sector indices in some markets experience significant impacts. More specifically, one sector index in France, four in Germany, and two in the US are affected by petroleum price fluctuations. Additionally, the equity returns of Consumer Staples and Information Technology sectors are influenced more frequently. In relation to other markets, no significant impacts on sector indices were obtained. Moya-Martinez et al. (2014) examine the responsiveness of fourteen Spanish sectors to fluctuations in petroleum prices, represented by the Brent grade, for the period from 1993 to 2010. The empirical results demonstrate that the effects of movements in petroleum prices on the Spanish sectors are quite limited, although the authors observe considerable variations from one sector to another. Xu (2015), using daily data, investigates the power of petroleum prices in predicting returns of the UK sector stock indices, represented by the FTSE All-Share industry indices, such as Basic Materials, Consumer Goods, Consumer Services, Financials, Health Care, Industrials, Oil and Gas, Technology, Telecom, and Utilities, over the period 1988 to 2013. The author found strong evidence of the relevance between movements in petroleum prices and the UK sector stock indices, where the effects on sector returns are heterogeneous. In particular, the results indicate that petroleum-related sectors are positively influenced, while the impact on petroleum-consuming sectors is adverse. Arouri (2011) utilised weekly data to investigate the reactions of twelve sector equity markets in Europe to petroleum price fluctuations over the period 1998–2010. The author's empirical results point to the existence of significant interconnections between movements in petroleum prices and the majority of sector equity markets. However, the strength of the observed link considerably varies across sectors. Furthermore, Arouri (2011) provides evidence of asymmetries in responses to petroleum price movements for some sectors. Scholtens and Yurtsever (2012) analyse the effects of petroleum prices on thirty-eight industries in the Euro area over the period 1983–2007. The authors applied the vector autoregression model and used various specifications of petroleum prices, such as changes in the price of petroleum, scaled petroleum prices, and net petroleum prices. Scholtens and Yurtsever (2012) documented that the effects considerably vary across sectors. Additionally, the significance of reported findings depends on petroleum price specifications.

Some works utilize a sample of Asian markets. Li et al. (2012) investigate the relationship between prices of petroleum and thirteen stock sector indices in China for the period 2001–2010. The authors obtained clear evidence indicating that, in the long run, the effects of increased real petroleum prices on Chinese sectoral equities are positive. Zhu et al. (2016), in their empirical study, which covers the period from March 1994 to June 2014, use the quantile regression technique to analyse the dependence between fluctuations in crude petroleum prices, represented by the WTI grade, and the equity returns of fourteen

Chinese industries. Their results show that the reactions of stock returns to petroleum price swings differ among industries. Furthermore, the authors found the presence of positive dependence, which is common at the lower quantiles across sectors, and suggest that strong dependence exists during bearish markets or recessions. Broadstock et al. (2014) examine the relationship between petroleum prices and energy-related stocks in the Asia-Pacific region, represented by six indices from four countries. The authors emphasise two channels of effects in their work, such as direct and indirect. The results indicate that the effects on stocks can be not only direct but also indirect through general market risk. They demonstrate that direct effects, although when considered only energy-related stocks, are not always present. However, indirect effects always exist for all indices considered in the study. Additionally, when the effects are significant, they are positive, implying that a sudden increase in the price of petroleum contributes to positive energy-related equity returns.

Other studies provide evidence from a global perspective. Nandha and Faff (2008) investigate the impact of fluctuations in the price of petroleum, represented by the WTI grade, on thirty-five global industry indices over the period 1983–2005. Their empirical results show that stock returns in all industries, with the exception of Mining and Oil and Gas, are negatively affected by movements in the price of petroleum. Furthermore, the authors suggest that interest rates and consumer confidence might be affected and depressed by higher petroleum prices, which consequently creates indirect channels to reflect elevated petroleum prices in stock prices. In addition, their findings demonstrate that increases and decreases in the price of petroleum exert a symmetric influence on the stock markets. Henriques and Sadorsky (2008) study interlinkages between petroleum prices, interest rates, technology, and alternative energy equity prices. The authors applied the four-variable vector autoregression model and employed weekly data for the period 2001–2007. Their empirical results show that petroleum prices have some power to explain the changes in equity prices of alternative energy firms. Additionally, the simulation findings demonstrate that a shock to petroleum prices has a little influence on alternative energy equity prices compared to a shock to technology equity prices. Ramos and Veiga (2011), using monthly data, studied the exposure of Oil and Gas sector indices from thirty-four countries to different factors, including movements in the price of petroleum, for the period 1998–2009. The authors found that the response of the Oil and Gas industry to petroleum price fluctuations in developed markets is much stronger than in their emerging counterparts. Furthermore, the reaction of Oil and Gas sector returns is asymmetric; that is, the impact of petroleum price increases in comparison to petroleum price decreases is greater. Kumar et al. (2012), through the application of the VAR model, examine the interconnection between petroleum prices and clean energy indices represented by the Wilder Hill New Energy Global Innovation Index, Wilder Hill Clean Energy Index, and S&P Global Clean Energy Index. The authors' empirical results indicate that the equity prices of clean energy companies are affected by petroleum prices. In particular, the effects of rising petroleum prices are positive. Cameron and Schnusenberg (2009) examine the relationship between petroleum prices and equity prices of automobile producers, namely General Motors (Detroit, MI, USA), Ford Motor Corp. (Dearborn, MI, USA), Daimler/Chrysler (Stuttgart, Germany), Toyota Motor Corp. (Toyota, Aichi, Japan), Honda Motor Co. (Minato, Tokyo, Japan), and Nissan Motors (Yokohama, Kanagawa, Japan), over the period 2001 to 2008. Their results show the existence of an inverse relationship between increasing prices of petroleum and automobile manufacturers' equity returns, where manufacturers of sport SUVs experience most of the impact. Nandha and Brooks (2009) investigate the role of petroleum prices in explaining the stock returns of transport sectors. Their analysis has a global perspective and includes thirty-eight countries. The authors used monthly data until July 2006, where start dates depended on the stock market history of each country. Nandha and Brooks (2009) classify countries into geographic and economic regions, namely Latin America, Europe, Asia Pacific, Developed, G7, and Emerging. Their empirical findings support the role of petroleum prices in determining equity returns in the transport sectors of countries within the Europe, G7, and Developed groupings, while no such evidence

is observed in countries falling within the remaining groupings. The empirical study of Kristjanpoller and Concha (2016) is concentrated on the analysis of the effects of fuel price movements, represented by the WTI grades of petroleum and Jet Fuel, on the stock returns of fifty-six airlines, which are part of the International Air Transport Association, for the period from 2008 to 2013. The results of their work show that airline equity prices are positively influenced by rises in fuel prices. The authors explain such findings in terms of asset prices, mainly commodities and equities, being positively associated with the bullish market and rising petroleum prices being a signal of higher economic growth.

5. Petroleum Price Shocks and Stock Markets

The empirical studies that have been reviewed in the previous sub-sections are mostly focused on the movements of petroleum prices when analysing the petroleum-equity market nexus. However, the identification of different sources that affect the price of petroleum is also crucial in order to better comprehend the interconnections between petroleum and stock markets.⁶ The price of petroleum, as with any commodity's price, is determined by market fundamentals, that is, by supply and demand (Kilian 2009). Kilian (2009) identified different types of supply and demand shocks in the crude petroleum market that explain changes in the price of petroleum, namely petroleum supply shocks, aggregate demand shocks, and petroleum-specific demand shocks. Additionally, Kilian and Park (2009) emphasise the importance of distinguishing between supply and demand shocks, as the empirical findings of studies that do not make such distinctions could be biased towards obtaining insignificant effects or not stable over time. Kilian and Park (2009), using monthly data, studied the effects of petroleum price shocks on aggregate and industry equity returns in the US over the period from 1973 to 2006. The authors show that the response of equity returns in the US to petroleum price shocks varies considerably, depending on the nature of movements in petroleum prices. In comparison to petroleum supply shocks, petroleum-specific demand and aggregate demand shocks are more important for explaining changes in US equity prices. Particularly, the effect of petroleum-specific demand shocks on equity returns is negative, whereas aggregate demand shocks positively influence equity returns. At the aggregate level, their findings demonstrate that, in the long run, nearly one-fifth of the variation in equity returns in the US is attributed to supply and demand shocks. At the industry level, the authors also document more significant reactions to petroleum demand shocks than to petroleum supply shocks, where the effects vary across sectors. In particular, equity returns in Precious Metals and Petroleum and Natural Gas industries positively react to aggregate demand shocks, but Automobile and Trucks and Retails industry returns respond negatively to petroleum-specific demand shocks. The empirical study of Kilian and Park (2009) is extended by Kang et al. (2016) through the distinction between shocks generated from US and non-US petroleum production and the analysis of their effects on both the US aggregate and sector equity returns for the period from 1973 to 2014. The authors found a positive association between equity returns and US petroleum production shocks. In addition, the findings indicate that petroleum supply and demand shocks, when the US and non-US petroleum production shocks are separated, have comparable importance in explaining equity returns. Furthermore, Kang et al. (2016) emphasise the importance of such disaggregation in order to understand the effects of petroleum supply shocks on the equity returns of sectors.

A large body of studies considers developed markets in their analyses. Apergis and Miller (2009) employ monthly data that cover the period from 1981 to 2007 to investigate the effects of structural petroleum market shocks on the equity markets of eight countries, namely Canada, the United States, France, the United Kingdom, Germany, Italy, Australia, and Japan. The authors construct three types of structural shocks, such as petroleum supply, global aggregate demand, and global petroleum demand. Their empirical findings indicate that the equity market returns of the studied countries do not significantly respond to shocks in the petroleum market; that is, the observed effects are small in magnitude. Kang and Ratti (2013) study the interrelationship between petroleum shocks, economic

policy uncertainty, and equity returns in the US at both aggregate and sectoral levels. The authors used monthly data for the period 1985–2011 and applied the structural VAR model. The impulse response function outputs demonstrate that the effects of petroleum supply shocks are not statistically significant for most of the period. Aggregate demand shocks positively affect US equity returns, while the impact of petroleum-specific demand shocks is negative. Additionally, the variance decomposition results demonstrate that, in the long term, approximately 32% of the variation in real equity returns is accounted for by structural petroleum shocks. With regard to responses from different sectors, the authors found that petroleum supply shocks positively influence returns of Automobiles and Trucks and Retail equities but depress equity returns in the Precious Metals industry. Aggregate demand shocks lead to statistically significant and persistent increases in equity returns in the Petroleum and Natural Gas sector, and petroleum-specific demand shocks negatively affect equity returns in the Automobiles and Trucks sector. The empirical work of Abhyankar et al. (2013) is focused on the analysis of the relationship between shocks that arise from petroleum prices and the Japanese equity market, represented by the Datastream Japan stock index. The authors applied the structural vector autoregressive model and employed monthly data covering the period from 1988 to 2009. Abhyankar et al. (2013) found that Japanese equity returns are positively, albeit not in a large magnitude, affected by petroleum supply shocks. Their results in relation to demand shocks confirm those of Kilian and Park (2009). The effect of aggregate demand shocks on stock market returns in Japan is positive. However, the reaction of the stock market to petroleum-specific demand shocks is negative. Guntner (2014) applies the methodology of Kilian and Park (2009) to analyse the effects of petroleum demand and supply shocks on the national equity markets of six OECD member countries, such as Canada, France, Germany, Japan, Norway, and the United States, for the period 1974–2011. The author found that petroleum supply shocks exercise little impact on the equity markets of all considered countries. Aggregate demand shocks cause increases in petroleum prices and equity returns, where the more persistent effect is reported for petroleum exporters, particularly for Norway. In addition, the results indicate that the effects of petroleum-specific demand shocks are positive for Norway and insignificant for Canada, while the impact on the stock markets of petroleum-importing countries is detrimental.

A group of more recent studies also focuses on advanced markets. Angelidis et al. (2015) investigate the ability of petroleum price shocks and volatility to forecast the state of equity market returns in the US over the period from 1989 to 2011. The authors follow the methodology of Kilian and Park (2009) in order to identify petroleum price shocks. Their empirical findings indicate that petroleum price shocks possess incremental power to predict the state of equity market volatility and returns in the US. More precisely, only petroleum-specific demand shocks have the power to forecast the state of equity market volatility. With regard to equity market returns regimes, only petroleum supply and aggregate demand shocks exercise significant effects. Bastianin et al. (2016) direct their attention towards examining the impact of three petroleum shocks on the level of stock market volatility in the G7 economies. The authors observe significant impacts of aggregate demand and petroleum-specific demand shocks on stock market volatility, albeit with differences in magnitude. On the contrary, responses to supply shocks are found to be insignificant. Sakaki (2019) studies the sensitivities of ten S&P 500 sector indices to petroleum shocks. The results indicate that equity returns in the majority of sectors are dampened by petroleum market-specific demand shocks, with the exception of the Energy and Utilities sectors, where the effects are positive. Petroleum supply shocks measured using shifts in US petroleum production in lieu of global petroleum production and aggregate demand shocks are found to induce positive reactions in stock sector returns. Utilising the nonlinear smooth transition VAR (STVAR) model, Hwang and Kim (2021) examine the influence of petroleum shocks on the US aggregate equity market and sector returns across the different business cycle stages. The authors observe that petroleum shocks exercise significant asymmetric effects on aggregate and sector equity returns over periods of economic contraction, where demand-side-driven petroleum shocks have larger im-

pacts in contrast to supply-side-driven petroleum shocks. In addition, the intensity and course of reactions are contingent on the nature of shocks and substantially differ among sectors.

Other works conduct investigations in the context of developing markets. Gupta and Modise (2013), applying the structural VAR model, studied the relationship between various petroleum price shocks and the stock market in South Africa for the period 1973–2011. The authors show that South African stock returns respond positively to aggregate demand shocks, whereas speculative demand shocks lead to declines in equity returns, which supports the findings of Kilian and Park (2009). However, the reaction of stock returns in South Africa to petroleum supply shocks is negative. The variance decomposition findings demonstrate that the contribution of petroleum supply shocks to the variability of South African equity returns is larger in contrast to other shocks. Fang and You (2014) investigate the impact of petroleum price shocks on the stock markets of three large emerging economies, namely China, India, and Russia, over the period from January 2001 to May 2012. The authors found mixed evidence of the effects of petroleum price shocks on equity prices. In particular, Indian equity returns respond positively to their own petroleum-specific demand shocks only in the first month, whereas global petroleum demand shocks negatively affect equity returns. In the case of Russia, stock returns are significantly decreased by global petroleum demand shocks, but Russian petroleum-specific supply shocks positively influence local stock returns. The impact of Chinese petroleum-specific demand shocks on regional equity returns is negative during some of the periods, while global demand shocks have no significant effects. Li et al. (2017) apply the structural vector autoregression model to investigate the impact of petroleum price shocks on the equity returns of listed firms in the petroleum industrial chain in China over the period from 2009 to 2014. The authors distinguish between four types of shocks: petroleum supply shocks, domestic demand shocks, global demand shocks and precautionary demand shocks. Their empirical findings indicate that the effects of petroleum supply shocks and precautionary demand shocks, compared to other shocks considered, are the most significant. Koh (2017), using the structural VAR model and monthly data, examines the relationship between various types of shocks and equity prices in fifteen Asian countries for the period from 1994 to 2014. The author confirms that the underlying causes of increases in the price of petroleum do matter. The empirical findings indicate that petroleum price increases caused by petroleum-specific demand shocks depress equity prices in the majority of cases, except for Malaysia, Korea, China, and Hong Kong. On the other hand, aggregate demand shocks lead to real equity return rises in most countries.

Other studies apply the innovative method of Ready (2018) for decomposing petroleum shocks that allows the utilisation of high-frequency data. Demirer et al. (2020) analyse the effects of petroleum shocks on equity market returns in twenty-one advanced and emerging economies using the daily frequency of data over the period from January 2000 to October 2018. The authors found that, regardless of countries' categorisation, equity market returns are positively impacted by petroleum demand shocks, while the consequences of petroleum supply shocks are more diverse. Wong (2020) investigates the sensitivities of Chinese sector equity returns to diverse types of petroleum shocks. The findings indicate that supply-driven shocks trigger positive equity returns in sectors, which can help ease concerns associated with supply constraints. On the other hand, the effects of demand shocks are found to be heterogeneous. Mishra and Mishra (2021) examine the responses of ten sector indices in India to petroleum demand, supply, and risk shocks for the period from 2010 to 2019. The authors document the positive and negative effects of demand and risk shocks, respectively, on all sector indices. With regard to supply shocks, only the reaction of the Metal sector index is observed to be significantly positive. Umar et al. (2020) focus on studying the static and dynamic connectedness between eight stock sector indices in Spain and three petroleum shocks. The static analysis revealed that the contribution of demand and risk shocks to fluctuations in sector stock returns is significant, whereas supply shocks exhibit negligible effects. Furthermore, the time-varying connectedness of demand and risk shocks with Spanish sector indices is greater, particularly during the

global financial crisis. Employing a sample of GCC and BRICS countries, Umar et al. (2021) explore the interrelationship between equity markets and petroleum shocks. Their findings indicate that the reactions of stock markets to shocks are country-specific, with demand and risk shocks also exerting a stronger impact on the level of connectedness.

Overall, the findings of the aforementioned studies can be summarised as follows: Petroleum demand shocks, irrespective of the methodology used to derive them, serve as an indicator of an economic expansion that, despite causing elevated petroleum prices, is viewed favourably by financial markets (Kilian and Park 2009; Ready 2018; Degiannakis et al. 2018), thereby credibly explaining positive reactions. The adverse impact of petroleum supply shocks and precautionary demand shocks is explained by the fact that the first generally leads to heightened inflationary pressures and economic contraction (Peersman and Van Robays 2012; Cashin et al. 2014), while the second transmits uncertainty in the petroleum market, caused by potential petroleum supply shortages as a result of various factors, to others (Degiannakis et al. 2018).

6. Time-Varying Linkages between Petroleum Prices and Stock Markets

Some recent empirical studies have documented the fact that the interrelationship between petroleum prices and stock markets changes over time and tends to be sensitive to major global events.⁷ A group of works focuses on single or multiple countries with developed markets. Reboredo (2010) uses the Markov-switching models to investigate the nonlinear effects of petroleum price shocks on stock markets in Germany, the Netherlands, the United Kingdom, and the United States over the period 1985–2006. The author provides evidence that equity market prices are differently affected by movements in petroleum prices. Specifically, the empirical findings indicate the presence of significant negative correlations at times of high uncertainty in equity markets, while at times of low uncertainty, the impact is null or minimal. Mollick and Assefa (2013) investigate the relationship between petroleum prices and equity returns in the US. The authors use daily data from 1999 to 2011 and consider three indices, such as the S&P 500, Nasdaq, Dow Jones, and Russell 2000. Their empirical findings indicate that the response of US equity returns to petroleum prices varies before, during, and after the global financial crisis. The association of equity returns with petroleum prices prior to and during the crisis is low. However, the effect on equity returns becomes statistically significant and positive during the latter period. Chang and Yu (2013), through the application of the MS-ARJI-GJR-GARCH-X model, examine the impact of crude petroleum price shocks, represented by the WTI grade, on equity returns in the US for the period 2001–2012. Their empirical outcomes show that petroleum price shocks exercise regime-dependent impacts, and behaviour differs during stable and turbulent periods. Ciner (2013) studied the impact of changes in the price of petroleum on the returns of aggregate stock markets and individual companies' equities from various industries in the US over the period 1986–2010. The author considers the time variation in the relationship and applies the frequency domain methods. The findings indicate that the linkage between petroleum and stocks is time-varying. Equity returns respond negatively to petroleum price shocks with a persistency of less than twelve months and greater than thirty-six months, while shocks with a persistency of twelve to thirty-six months cause positive reactions. Furthermore, the results are mostly robust when individual equity returns are employed, with the exception of petroleum firms' equities that have positive associations with increases in petroleum prices. Degiannakis et al. (2013) examined the time-varying relationship between fluctuations in the price of petroleum and the returns of ten European sector indices for the period 1992–2010. The authors considered various origins of petroleum price shocks in their study. Their empirical findings indicate that the interrelationship between petroleum price movements and sector indices changes over time and differs among industries. In addition, precautionary demand shocks result in nearly zero correlation levels, petroleum supply shocks generate positive correlations with low to moderate levels, and aggregate demand shocks lead to substantial variations, either positive or negative, in correlation levels. Reboredo and Rivera-Castro (2014), through

the wavelet multi-resolution analysis, study the intrarelationship between petroleum and equity markets in the US and Europe for the period from June 2000 to July 2011. The authors conduct both market- and sectoral-level analyses. They found that aggregate and sector equity market returns were not affected by the petroleum price movements, except for stocks of oil and gas firms, in the period preceding the financial crisis. However, with the commencement of the financial crisis, the positive interdependence and contagion between petroleum and equity markets are evident at both aggregate and sectoral levels. The empirical work of Martin-Barragan et al. (2015) focuses on the analysis of the influences of equity market crashes and petroleum shocks on correlations between petroleum and equity markets in the United States, Germany, the United Kingdom, and Japan for the period from February 1990 to November 2011. The obtained findings show that in non-crisis periods, the correlation between equity and petroleum markets has a tendency to be stable. On the contrary, during financial and petroleum shocks, this correlation fluctuates at lower and higher frequencies.

Other works report evidence in the context of emerging markets. Bhar and Nikolova (2010) study the relationship between petroleum price changes and the Russian stock market, represented by the AK&M Composite index, for the period 1995–2007. The authors found that Russian stock market returns and volatility are significantly affected by movements in petroleum prices. In addition, the empirical results indicate that the conditional correlation between petroleum prices and stock returns in Russia is negative after the unstable global events that took place in 2001, 2003, and 2006. Mohanty et al. (2010) analyse the relationship between the prices of petroleum and the equity returns of oil and gas companies from Central and Eastern European countries for the period from 1998 to 2010. Their overall empirical findings indicate the absence of a significant association between petroleum prices and the stock values of oil and gas firms. In contrast, the results obtained from the sub-period analysis suggest that the equity returns of some oil and gas firms are affected by petroleum prices, where the levels of exposure vary across companies and over time. Broadstock et al. (2012), through the employment of weekly data, examine the relationship between prices of petroleum and equities in the Chinese energy sector over the period from 2000 to 2011. Their results suggest that the correlation between changes in petroleum prices and energy-related equity returns is time-varying. In particular, the interrelation between these variables strengthened following the global financial crisis.

Several studies consider a combination of developed and emerging markets. Broadstock and Filis (2014) use monthly data to examine time-varying correlations between three types of petroleum price shocks and equity market returns for the period 1995–2013. The authors employ a sample of both aggregate stock market and sector indices from China and the US. The empirical findings indicate that stock market responses to petroleum price shocks fluctuate over time. Specifically, the Chinese equity market demonstrates more resilience to petroleum price shocks as opposed to the equity market in the US. Furthermore, the authors report that shocks, depending on their origin, exercise diverse effects on stock markets, which substantially vary across sectors. Reboredo and Ugolini (2016) examine the effects of petroleum price fluctuations on various equity return quantiles in three developed and BRICS countries for the period from 2000 to 2014. Their empirical results indicate the existence of asymmetric effects that are limited in the period prior to the financial crisis but strengthen after the commencement of the financial crisis. Zhang (2017) investigates the time-varying correlations between the price of petroleum and equity markets in China and the US, represented by the CSI 300 and S&P 500 indices, respectively. The author used daily data for the period from 2002 to 2013 and applied the mixed asymmetry dynamic conditional correlation model in their study. They found that the correlation structure was significantly changed by the great shocks that occurred in 2003 and 2008. For the US, the correlation between the equity market and petroleum prices changes from positive to negative after the great shock of 2003, while the correlation increases in the case of China. Furthermore, the global financial crisis causes upsurges in correlations between petroleum prices and the equity markets of both countries. Zhu et al. (2017) apply the

two-stage Markov regime-switching approach to study the influence of petroleum demand and supply shocks on equity returns in ten countries for the period 1997–2015. Their empirical outcomes show that in the low-volatility regime, equity returns are minimally impacted by structural petroleum shocks. However, the effects in the high-volatility regime are statistically significant. Furthermore, the authors observe that aggregate demand and petroleum-specific demand shocks significantly influence equity returns in comparison to petroleum supply shocks. More specifically, equity returns are increased by positive aggregate demand shocks, while positive petroleum-specific demand shocks lead to decreases in equity returns.

7. Petroleum Price and Stock Market Volatility

The strengthened financialisation of commodity markets along with advancements in technology contributed to the spread of announcements, news, speculative rumours, and investors' sentiment between petroleum and stock markets across the world, particularly given the nearly continuous operation of the former. In this connection, a number of empirical studies have examined the interrelationship between the volatility of the petroleum and equity markets.⁸ Some of these works concentrate on developed markets. Aloui and Jammazi (2009) apply the two-regime MS-EGARCH model to study the effects of crude petroleum volatility shocks on the behaviour of equity markets in France, Japan, and the UK. The authors use monthly data from the CAC 40, Nikkei 225, and FTSE 100 indices over the period from 1989 to 2007. Their empirical findings provide evidence of the significant role of petroleum price increases in determining equity return volatility and the likelihood of transition between regimes. Vo (2011) utilises the multivariate stochastic volatility framework to investigate the association between volatilities in the petroleum and US equity markets. The author documented the presence of volatility interdependencies between two markets. Mensi et al. (2013) employ the VAR-GARCH model to study the volatility linkages between the S&P 500 index and different types of commodity indices, such as gold, wheat, beverage, and petroleum, represented by both the WTI and Brent grades, over the period from 2000 to 2011. Their empirical findings provide evidence of the bidirectional volatility transmission between the S&P 500 index and the WTI grade of petroleum. However, the results differ in the case of the Brent grade. More precisely, equity market volatility is significantly affected by Brent crude petroleum volatility and not the reverse. The empirical study of Chang et al. (2013) is focused on the analysis of volatility spillovers between petroleum returns, represented by spot, futures, and forward prices of the WTI and Brent grades, and returns of the NYSE, Dow Jones, FTSE 100, and S&P 500 indices for the period 1998–2009. The authors used daily data and applied different types of multivariate GARCH models. Surprisingly, they found little evidence of volatility transmissions between the prices of petroleum and equities. Kang et al. (2015) apply the structural VAR model to examine the influence of petroleum price shocks on US equity market volatility for the period from 1973 to 2013. The authors also focus on three volatility measures, namely realised, conditional, and implied volatility. Their empirical results show that petroleum aggregate demand shocks negatively affect all measures of volatility. Furthermore, the association of petroleum-specific demand shocks with each of the volatility measures has been negative over the first several months. The effects of petroleum supply shocks are associated with significant increases in implied volatility, whereas the effects on conditional and realised volatility are not significant. Salisu and Oloko (2015) investigate interactions between the Brent and WTI grades of petroleum and the US stock market, represented by the S&P 500 index, over the period 2002 to 2014. The authors account for structural breaks in their study, the dates of which coincided with the global financial crisis period. Their findings indicate that during the period preceding the structural break, the volatility spillover from the petroleum market to the US equity market was significant for WTI. However, the volatility transmission from the petroleum market to the stock market heightened after the global financial crisis in the case of both the Brent and WTI grades. Khalfaoui et al. (2015) use multivariate GARCH models and wavelet analysis

to study the volatility spillover effects between petroleum prices and equity markets of the G7 countries, namely Canada (S&P/TSX), France (CAC 40), Germany (DAX), Italy (FTSE MIB), Japan (Nikkei 225), the United Kingdom (FTSE 100), and the United States (S&P 500), for the period from 2003 to 2012. The authors provide evidence of significant volatility transmissions between the petroleum and equity markets. Ewing and Malik (2016) employ bivariate and univariate GARCH models to investigate spillovers of volatility between the WTI grade of petroleum and the US equity market over the period from 1996 to 2013. The authors incorporate structural breaks into their study. They found no evidence of volatility transmissions between two markets when structural breaks were disregarded. However, after incorporating structural breaks, the authors detect evidence of volatility spillovers between petroleum prices and the US equity market. Utilising daily data over the period from 21 January 2020 to 2 July 2020, Adekoya and Oliyide (2021) examine the role of COVID-19 in driving the connectedness between commodities (crude petroleum and gold) and financial markets (the S&P 500 index, dollar exchange rate, and bitcoin). Their findings indicate that the total time-varying volatility connectedness remained high during the first half of the estimation period and subsequently descended to a lower level. Among the studied assets, crude petroleum acts as the major net transmitter for others.

Other studies consider both developed and emerging markets. Malik and Ham-moudeh (2007) study the volatility transmission interrelationship between petroleum market and equity markets in the US, Bahrain, Kuwait, and Saudi Arabia over the period 1994–2001. Their empirical results indicate that volatility spills over from the petroleum market to the equity markets of Bahrain, Kuwait, and Saudi Arabia. However, the authors found that the petroleum market is affected by the equity market volatility in the case of Saudi Arabia, thereby underscoring the important role of the country in the global petroleum market. Maghyereh et al. (2016) examined interconnections between implied volatilities of petroleum and stock market indices of eleven countries, namely Canada, Germany, India, Japan, Mexico, Russia, South Africa, Sweden, Switzerland, the UK, and the US, for the period 2008–2015. Their findings show that the interrelationship between the volatilities of markets is established by bidirectional information transmissions. More specifically, the information spillover is greater from petroleum to stock markets than in the opposite direction. Moreover, the authors detect large transmissions during the period of global recovery, from the middle of 2009 to 2012. Tiwari et al. (2021) analyse the connectedness of volatility between petroleum, represented by Brent and WTI grades, and stock markets in a set of developed and emerging economies. They found that petroleum, irrespective of the grade, is the net recipient of volatility spillovers. Furthermore, volatility interactions are dynamic in nature and have been observed to be impacted by major global events. Utilising a sample of seven advanced and five emerging economies, Mensi et al. (2021) investigate volatility transmissions between stock markets and strategic commodities such as petroleum and gold. The authors document evidence of the dynamic volatility spillovers that strengthen during major events. The stock markets transmit a greater level of volatility to petroleum than they receive on the short-term horizon. The comparable results are detected over intermediate- and long-term horizons, aside from the stock markets of China, Japan, and Russia, where more spillovers in the reverse direction are observed.

A collection of studies considers emerging markets. Arouri et al. (2011a) examine the transmission of volatility between petroleum prices, represented by the Brent grade, and the equity markets of the Gulf Cooperation Council countries, namely Bahrain, Kuwait, Oman, Saudi Arabia, Qatar, and the UAE. The authors applied the VAR-GARCH model and used daily data covering the period from June 2005 to February 2010. Overall, the authors provide evidence of the existence of significant volatility spillovers, particularly during the crisis sub-period, between the prices of petroleum and the equity markets of the studied countries. Awartani and Maghyereh (2013) also examine the volatility spillover effects between the petroleum market and equity markets of the Gulf Cooperation Council countries, but for the period from 2004 to 2012. Their empirical results demonstrate that bidirectional volatility transmissions, although asymmetric, exist between the petroleum

market and equity markets. More specifically, volatility spillovers are more pronounced from the petroleum market to equity markets than in the reverse direction. Furthermore, such trends were evident during the financial crisis in 2008 and strengthened afterwards. The study by Lin et al. (2014) focuses on developing markets in the African region. The authors applied DCC-GARCH, VAR-GARCH, and VAR-AGARCH models to examine volatility transmissions between the petroleum market and equity markets in Ghana and Nigeria, represented by the GSE All-Share and NSE All-Share indices, for the period 2000–2010. Their empirical findings provide evidence of significant volatility spillovers between petroleum and equity market returns. The documented effects are stronger in the case of Nigeria, while for Ghana, the volatility transmission is more evident from the petroleum market to the equity market. Yousaf and Hassan (2019) investigate volatility interactions between petroleum stock markets in nine Asian emerging economies. The authors found that volatility spillovers are contingent on study periods and differ across stock markets. Specifically, throughout the full sample and the subprime meltdown in the United States, volatility transmissions are predominantly unilateral from petroleum to stock markets, while during the turbulent period in the Chinese stock market, volatility spillovers are bilateral between petroleum and Indian and Korean stock markets. Focusing on the GCC stock markets, energy, and precious metal commodities, Al-Yahyaee et al. (2019) observe that petroleum, represented by the WTI grade, is the net contributor of volatility. Furthermore, the total volatility spillover index is found to intensify at times of economic turmoil and geopolitical tensions. Sarwar et al. (2020) study the spillover of volatility between petroleum and equity market returns in China, India, and Pakistan. Their findings, which vary across three subperiods and frequencies of data, indicate that volatility transmissions are mixed for India, bilateral for Pakistan, and unilateral for China.

Several works stress the importance of a sectoral examination. Hammoudeh et al. (2004), using daily data, analysed the effects of crude petroleum spot and futures prices on five equity indices of the US petroleum sector for the period from July 1995 to October 2001. The authors report bidirectional volatility interactions between petroleum sector indices and petroleum spot and futures prices. Employing weekly data for the period from 1992 to 2008 and the bivariate BEKK-GARCH model, Malik and Ewing (2009) investigate the volatility transmission mechanism between prices of petroleum and five sector indices in the US. Their findings provide evidence of the substantial transmission of volatility between petroleum prices and sector indices. Nevertheless, such effects are not detected in the Financial and Industrial sectors. Elyasiani et al. (2011) study the effects of shifts in petroleum futures returns and volatility on the equity returns of the thirteen US sectors for the period 1998–2006. The authors categorise these sectors into four groups: petroleum-substitute, petroleum-related, petroleum-users, and financials. Their empirical findings indicate that the excess returns of petroleum-user sectors are influenced by variations in the volatility of petroleum futures returns. On the other hand, sectors in the petroleum-related and petroleum-substitute groups are more affected by fluctuations in petroleum futures returns. Sectors in the financial group are impacted by both movements in petroleum futures returns and volatility. Arouri et al. (2011b) provide evidence of bidirectional volatility spillovers between stock sectors of the United States and petroleum and mostly unidirectional volatility spillovers from petroleum to European stock sectors.

Following the same line, Arouri et al. (2012) apply the VAR-GARCH model to investigate volatility spillover effects between prices of Brent crude petroleum and European equity markets, represented by the DJ Stoxx Europe 600 aggregate index and seven sector indices, over the period 1998 to 2009. Their empirical findings demonstrate the significant transmission of volatility between prices of petroleum and sector equity returns in Europe, where the effects are more apparent from petroleum to sector equity markets. Furthermore, the results indicate that volatility interactions differ across sectors. Sadorsky (2012) examines volatility spillovers and correlations between prices of petroleum and equity prices of technology and clean energy firms for the period 2001–2010. The authors applied multivariate GARCH models and used daily data from the NYSE Arca Technology and WilderHill

Clean Energy indices. The findings indicate that the dynamic conditional correlation between the equity prices of clean energy and technology firms is higher than between clean energy equity prices and petroleum prices. In addition, the results obtained from the DCC type of model show the presence of statistically significant volatility transmission from petroleum prices to equity prices of clean energy firms. Degiannakis et al. (2014) investigate the effects of petroleum price shocks on the volatility of equity markets, represented by the Euro Stoxx 50 and ten Dow Jones sector indices, over the period 1999 to 2010. The authors in their study consider the realised, conditional, and implied volatility measures. Their empirical findings show that volatilities of the aggregate equity market and sectors are not influenced by petroleum supply and petroleum-specific demand shocks, while aggregate demand petroleum shocks result in a reduction of volatilities. Alsaman (2016) examines the effects of petroleum price uncertainty on equity returns in the US at both aggregate and sector levels. The author applied the bivariate GARCH-in-mean VAR model and used monthly data covering the period from 1973 to 2014. The author's results show no evidence of statistically significant impacts of petroleum price volatility on stock returns in the US. Alsaman (2016) suggests that the application of hedging strategies in order to minimise risks arising from petroleum price fluctuations might explain such findings. Applying the VAR-BEKK-GARCH model, Belhassine (2020) documents the non-uniform and dynamic nature of volatility transmissions between petroleum and most of the Eurozone equity sector indices. Costola and Lorusso (2022) scrutinise volatility interactions between energy commodities and six Russian sector indices over the period from January 2005 to December 2020. The authors detect that the magnitude of volatility spillovers is greater from sector indices to petroleum than in the opposite course. Among sectors, Oil and Gas and Metals and Mining are the largest contributors of volatility to petroleum. Furthermore, the total dynamic volatility connectedness is found to be sensitive to global events, particularly the financial crisis and the recent pandemic.

8. Petroleum Prices and Stock Markets in Petroleum Exporting and Importing Countries

It is generally expected that petroleum price changes would have different effects on stock markets in net petroleum exporting countries compared to net petroleum-importing countries (Smyth and Narayan 2018). More specifically, in the case of petroleum-importing countries, a surge in the petroleum price generally leads to higher levels of production costs, given that crude petroleum and petroleum-related products are among the fundamental components in the manufacturing process, while for petroleum exporting countries, it triggers an influx of wealth.⁹ Several empirical studies have analysed how the relationship between petroleum prices and equity returns differs across petroleum exporting and importing countries. Park and Ratti (2008) examine the effects of petroleum price shocks and volatility on the equity returns of the United States and thirteen European countries over the period from 1986 to 2005. Their empirical findings indicate that the impact of petroleum price shocks on equity returns in the studied countries is statistically significant. More precisely, the effects of shocks are positive in the case of Norway, a petroleum exporter, and negative for the remaining European countries and the United States. Furthermore, equity returns in most European countries are significantly depressed by the amplified volatility of petroleum prices. Ramos and Veiga (2013) use a sample of thirteen petroleum-importing and five petroleum-exporting countries covering the period 1988–2009 to investigate the impact of petroleum prices on stock markets. Their results show that the effects of petroleum price changes for petroleum exporters and importers run on different routes. Specifically, petroleum price hikes positively affect stock markets in petroleum-exporting countries, while the impact on the stock markets of petroleum-importing countries is negative. Petroleum price drops negatively influence the stock markets of two country groups, while the effect is larger for petroleum exporters. Furthermore, the results show that the impact of petroleum price volatility on the stock markets of petroleum exporting countries is positive and on those of petroleum-importing countries is

negative. The empirical study of Wang et al. (2013) is focused on the analysis of the effects of petroleum price shocks on stock market returns in seven petroleum exporting and nine petroleum importing countries for the period 1999–2011. The authors employ the structural VAR model in their work. The main findings indicate that the sign, magnitude, and duration of reactions of equity market returns to petroleum price shocks are contingent on the origin of shocks and the position of countries in the petroleum market, that is, whether they are net exporters or importers. In addition, the results show that petroleum price shocks have more power to explain variations in stock markets for petroleum-exporting countries in comparison to petroleum-importing countries.

More recently, Le and Chang (2015) examined the effects of fluctuations in Dubai crude petroleum prices on equity returns for the period from January 1997 to July 2013. The authors focus on petroleum exporting, importing, and refining economies represented by Malaysia, Japan, and Singapore, respectively. The main estimation period in their study is split into three sub-periods. Overall, the empirical findings indicate that the responses of stock markets to petroleum price hikes vary across different countries and periods. In the first sub-period, the immediate effect of petroleum price shocks on the Malaysian stock market is positive but turns negative in the long run. The impact in the case of Singapore is positive and gradually becomes insignificant. The Japanese stock market generally has positive reactions. In the second sub-period, the instant responses of stock markets in all nations are positive. The Japanese and Singaporean stock markets experience unfavourable influences in the third sub-period, while for Malaysia, the impact is positive only for a short duration and then changes to negative. Employing a sample of nine exporters and seven importers, Wang and Liu (2016) provide evidence of country-specific volatility spillovers. More precisely, volatility spills over from the stock markets of three importers, Germany, the UK, and the US, to the petroleum market, while in the case of exporters, the stock market volatility of Canada, Norway, Russia, and Venezuela is affected by the petroleum market volatility. Employing the nonlinear Panel ARDL approach, Salisu and Isah (2017) examine the asymmetric responses of stock prices to petroleum price changes in five net petroleum exporting and eight net petroleum importing countries over the period from 2000 to 2015. Their empirical findings indicate that the reactions of stock prices in petroleum exporting and importing countries to movements in the price of petroleum are asymmetric, although stronger responses were detected in the former group. Ashfaq et al. (2019) observe volatility spillovers between the petroleum and equity markets of one importer (South Korea) and two exporters (Saudi Arabia and Iraq). The authors found no evidence of significant effects in the cases of India and Japan (importers) and the United Arab Emirates (exporter). Belhassine and Karamti (2021) report the presence of significant volatility transmissions between the petroleum and stock markets of exporters and importers, which vary across considered markets and time scales. Enwereuzoh et al. (2021) focus on equity markets in African exporting and importing economies in order to examine their sensitivities to petroleum shocks for the period 2000–2018. Their findings demonstrate precautionary demand shocks exert significant effects on most of the studied equity markets, and aggregate demand shocks predominantly impact the equity markets of exporting countries. However, the authors obtain scant evidence that petroleum supply shocks affect the stock markets of exporting and importing countries. The study of Akyildirim et al. (2022), which focuses on examining the connectedness between MSCI energy stock indices of petroleum exporters and importers, shows that the first group of countries are mainly transmitters of shocks, while the latter act as receivers of shocks.

Several studies centre around dynamic correlations. Filis et al. (2011), in their work, utilise the sample of three petroleum exporting and importing countries to study time-varying correlations between prices of petroleum and stock markets for the period from 1987 to 2009 based on the DCC-GARCH-GJR model. The authors observe that while time-varying correlations of petroleum and equity prices do not differ for petroleum exporters and importers, they depend on sources of petroleum price shocks. Thus, aggregate demand-

side shocks cause positive and precautionary demand shocks negative correlations between markets, but supply-side shocks have no influence on the relationship. Antonakakis and Filis (2013) use the DCC-GARCH model to examine the effects of changes in petroleum prices on the time-varying correlation between stock markets. The authors consider in their study stock market indices from two petroleum-exporting and three petroleum-importing nations. Their findings show that petroleum price movements influence correlations between the stock markets of petroleum-importing countries, while no significant effects on correlations are detected in the case of petroleum-exporting economies. Applying the GJR-DCC-GARCH model, Guesmi and Fattoum (2014) investigate co-movements between Brent crude petroleum prices and stock market indices of four petroleum-exporting economies and five petroleum importing economies over the period from 2000 to 2010. The authors' findings show no differences in dynamic correlations for petroleum-exporting and importing countries. Petroleum demand shocks significantly affect the relationship between petroleum prices and stock markets, particularly in exporting economies. Petroleum supply shocks exert an impact on a correlation only in exporting countries. Boldanov et al. (2016) investigate the relationship between petroleum price and equity market volatilities in petroleum exporting, Canada, Norway, and Russia, and importing, China, Japan, and the US, countries for the period from 2000 to 2014. Their findings indicate that correlations between volatilities are country-specific and vary over time. Particularly, the relationship between petroleum and stock market volatilities is mostly positive in the case of petroleum-importing nations, while more negative correlations are reported for petroleum-exporting nations. In addition, the authors show that correlations are sensitive to different global events. Mokni (2020) analyses the dynamic responses of stock markets in exporting and importing countries to aggregate demand, petroleum supply, and petroleum-specific demand shocks. The author observes that the impact of aggregate demand shocks is positive in most cases, while the effects of petroleum supply shocks are minimal and negative. In addition, the stock markets of exporting countries react positively to petroleum-specific demand shocks, but their counterparts from importing countries react negatively, except for China.

Some works focus exclusively on petroleum-importing nations. The study by Masih et al. (2011), which considers South Korea, one of the largest petroleum importers, shows that movements in petroleum prices significantly affect the country's stock market. Specifically, reactions of the South Korean stock market to petroleum price shock and volatility are negative, with a more profound effect in the latter case. Cunado and Perez de Gracia (2014) analyse the impact of petroleum price shocks on equity returns in twelve petroleum-importing European countries. The authors employed monthly data covering the period from 1973 to 2011. Their empirical results indicate that the effects of fluctuations in petroleum prices on stock market returns in the majority of countries are significant and negative. Furthermore, the authors found that petroleum supply shocks, in comparison to petroleum demand shocks, have larger negative impacts on equity market returns when both world and local petroleum prices are employed. Petroleum demand shocks negatively influence the returns of stock markets only in five countries and positively affect equity returns in France and Denmark when national petroleum prices are used. Bouri (2015) investigates the transmission of petroleum price volatility to the equity markets of four petroleum-importing countries from the MENA region, namely Jordan, Lebanon, Morocco, and Tunisia, for the period 2003–2013. In the pre-crisis period, the author found no evidence of volatility transmissions between the global petroleum market and stock markets in all countries. However, different findings are reported for the post-crisis period. While stock markets show no responses to the petroleum price volatility in the case of Lebanon and Morocco, the bidirectional volatility spillover effects are detected between the world petroleum market and the equity market in Jordan. Additionally, for Tunisia, the unidirectional transmission, although weak, of petroleum volatility to the stock market is present. Silvapulle et al. (2017) applied the nonparametric and parametric panel data approaches to study the relationship between prices of petroleum and stock market in-

dices of large petroleum-importing nations, namely China, France, Germany, India, Italy, Japan, Singapore, South Korea, Spain, and the United States, for the period 1999 to 2015. The reported findings of both methods indicate that the effects of petroleum prices on stock market indices are mostly positive, although they vary during different periods. Overall, the authors conclude that the nonparametric method better captures the dynamic petroleum-stock market relationship in comparison to the parametric method. Sarwar et al. (2019) investigate volatility spillovers between the petroleum market and equity markets of the largest petroleum importers in Asia (China, Japan, and India) over the period from 2000 to 2016. Their results indicate that volatility spillovers are bidirectional between petroleum and the Japanese stock market and unidirectional from the Indian stock market to petroleum, but absent in the case of the Chinese stock market.

Other studies consider petroleum-exporting nations. Employing the structural VAR model, Bjornland (2009) analyses the influence of petroleum price shocks on equity returns in Norway, a net petroleum exporting country. The author's findings suggest that higher petroleum prices increase equity returns in the country. More specifically, the results demonstrate that following an increase of 10% in the petroleum price, equity returns rise instantly by 2.5%. The highest effect is detected after 14–15 months, which gradually fades. Mohanty et al. (2011) investigate the relationship between petroleum price fluctuations and stock market returns in Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, and the UAE for the period 2005–2009. Their empirical results show that declines in the price of petroleum negatively impact equity market returns in all the studied countries. In contrast, petroleum price increases have significant and positive effects on the stock market returns of only two countries: Saudi Arabia and the UAE. Furthermore, the industry-level analysis conducted for four countries demonstrates that reactions to petroleum price shocks considerably vary across considered industries. Arouri and Rault (2012) examine long-term links between the prices of petroleum and the stock markets of the four Gulf Cooperation Council countries, namely Bahrain, Kuwait, Oman, and Saudi Arabia. The authors applied seemingly unrelated regression methods, bootstrap panel cointegration techniques, and employed monthly data for the period from 1996 to 2007. Their results provide evidence for cointegration between petroleum prices and equity markets. In addition, the empirical findings indicate that rises in the price of petroleum positively affect stock markets in all countries, except for Saudi Arabia. Demirer et al. (2015) explore the effects of petroleum price risk exposure on equity returns in net petroleum-exporting nations. The authors focus on the Gulf Cooperation Council countries and use firm-level data covering the period from 2004 to 2013. The results demonstrate that equities that are more responsive to movements in the price of petroleum produce notably higher returns. Thus, the obtained findings suggest that exposure to petroleum prices can act as a forecaster of equity returns in the studied countries. Gil-Alana and Yaya (2014) analysed the relationship between monthly Brent crude petroleum prices and the Nigerian stock market based on fractional integration and cointegration approaches over the period from 2000 to 2011. The authors document evidence of the positive relationship between the studied variables, although with a short-term effect that remains significant only for the first three months. Basher et al. (2018) use monthly data to study the impact of different petroleum shocks on the performance of stock markets in eight petroleum exporting nations, namely Canada, Mexico, the UK, Norway, Russia, Saudi Arabia, Kuwait, and the UAE. Their results show that magnitudes and signs of petroleum shocks vary among the studied countries. The impacts of petroleum demand shocks on equity returns in all countries, with the exception of Mexico and the UK, are statistically significant. Petroleum supply shocks exert statistically significant effects in the case of Kuwait, the UAE, and the UK. The stock returns in Russia, Norway, Saudi Arabia, Kuwait, and the UAE are influenced by idiosyncratic petroleum market shocks. Petroleum inventory shocks affect equity returns in Russia, Canada, Kuwait, and the UAE. Lastly, petroleum market shocks have no effect on equity returns in Mexico. The authors explain the findings obtained by the absence of large publicly traded oil and gas corporations in the country.

A few studies conduct sectoral investigations. Bouri et al. (2016) examine the association between petroleum prices and returns of Financials, Industrials, and Services stock sectors in Jordan, a small petroleum-importing country from the MENA region. Their findings show that the returns of the Financials and Services sectors are significantly influenced by petroleum shocks, whereas the effect is insignificant in the case of the Industrial sector. The reported impact is more noticeable during the period that follows the Uprising in the region. Furthermore, the authors found evidence of risk transmission from the petroleum market to the Industrial sector only. Bagirov and Mateus (2022) scrutinise volatility transmissions between petroleum prices and manually constructed stock sector indices of Mexico (net exporter) and the United Kingdom (net importer) over the period from January 2005 to September 2018. The authors document the presence of significant bilateral volatility spillovers, which are more evident for sectors of Mexico. In the case of the United Kingdom, the volatility cross-effects are observed between petroleum and only two sectors.

9. Petroleum Prices and Firm-Level Stock Returns

A group of studies have analysed the relationship between petroleum prices and equity returns employing firm-level data.¹⁰ Sadorsky (2008) investigates the relationship between fluctuations in petroleum prices, equity prices, and firm size using data on 1483 companies from the S&P 1500 index, which comprises small, medium, and large companies, covering a period of seventeen years. The author observes that the effects of movements in the price of petroleum on equity returns are asymmetric and vary with the size of companies, which are stronger in the case of medium-sized companies. Petroleum price rises exercise a greater impact on equity returns in comparison to petroleum price decreases. Furthermore, the findings indicate that increases in petroleum prices lead to drops in company-level equity returns. Narayan and Sharma (2011) use daily data to examine the impact of petroleum prices on the stock returns of 560 US companies from fourteen sectors listed on the New York Stock Exchange for the period 2000–2008. Their main results suggest that the responses of companies to petroleum price movements depend on their size and sectoral belonging. More specifically, petroleum price increases lead to rises in returns for companies in the energy and transportation sectors, while corporations in the remaining sectors experience decreases in returns, which could be attributed to different degrees of dependence on petroleum. In addition, the authors found that the relationship between petroleum prices and returns of small-sized companies in the majority of instances is positive and statistically significant, but it turns out to be negative as firms' sizes grow, thereby pointing to a greater flexibility that small-sized companies may have in terms of efficiently adapting to changing market conditions. Dayanandan and Donker (2011) explore the interrelationship between petroleum prices, firm size, financial performance, and capital structure of oil and gas companies in the North American region over the period from 1990 to 2008. The authors used in their study data on 200 large corporations listed on the US stock exchange. Dayanandan and Donker (2011) found that the impact of petroleum prices on measures of the financial performance of oil and gas companies is significant and positive. In addition, their empirical findings indicate that the global financial crisis negatively affects the prices of petroleum and the performance of companies.

Aggarwal et al. (2012) analysed the influence of changes in petroleum prices on various firm characteristics of seventy-one companies from the S&P Transportation sector index over the period 1986–2008. The authors found asymmetrical exposures to petroleum price shifts. In addition, their results indicate that petroleum price upsurges negatively influence the returns of transportation companies, while declines in petroleum prices increase risks. The divergence in findings of Narayan and Sharma (2011) and Aggarwal et al. (2012) related to transportation stocks could be attributed to different empirical methodologies employed. Mohanty et al. (2013) study the effects of petroleum price fluctuations on equity returns and other firm-specific characteristics of fifty-four companies from the US oil and gas industry for the period from 1986 to 2008. Their results show that both petroleum price increases and decreases influence the returns of oil and gas companies. However, the reactions

of equity returns associated with negative petroleum price movements are significantly greater, and equity risks are more affected by declines in the price of petroleum. In addition, the authors found that responses to fluctuations in petroleum prices are contingent on the characteristics of companies and differ across sub-sectors. Phan et al. (2015) examined the impact of movements in petroleum prices on the equity returns of petroleum-consuming and producing companies in the US over the period from 1986 to 2010. The authors found evidence indicating that the influence of petroleum price increases on the equity returns of petroleum producers is positive, whereas for petroleum consumers, the negative impact is documented. In addition, their results show that the magnitude of observed effects differs across sub-sectors of petroleum producers and consumers. Gupta (2016) used company-level monthly data covering seventy countries to examine the effects of petroleum price shocks, competition, and country-level determinants on the equity returns of oil and gas companies over the period from 1983 to 2014. The author's findings indicate that petroleum price shocks positively affect company-level returns. Oil and gas firms from petroleum-producing countries showed higher sensitivity to market stress and petroleum price fluctuations. Furthermore, companies located in a non-competitive environment are less influenced by decreases in the price of petroleum. The empirical work of Bagirov and Mateus (2019) is focused on the analysis of the effects of crude petroleum price movements on the financial performance of 137 listed and 531 unlisted oil and gas corporations from the Western European region for the period from 2005 to 2014. Their results demonstrate that the performance of listed oil and gas corporations is positively affected by crude petroleum prices. Furthermore, the recent geopolitical crisis exerted negative effects on listed and unlisted corporations, while the global financial crisis adversely affected only listed corporations.

Narayan and Sharma (2014) analyse the impact of petroleum prices on the equity return volatility of 560 companies from fourteen sectors listed on the New York Stock Exchange over the period from 2000 to 2008. The authors' results indicate that the effects of petroleum prices vary depending on the sectors and sizes of companies. In particular, petroleum price rises lead to equity return volatility increases for the banking sector and decreases in equity return volatility for all other sectors. Tsai (2015) investigates the reactions of equity returns in the US to petroleum price shocks for the period from 1990 to 2012. The author employs the daily data of 682 companies. The empirical findings indicate that the equity returns of companies were negatively influenced prior to the global financial crisis. However, the effects are positive during and after the crisis periods. In addition, the author observes heterogeneous responses across industries. Particularly, the reactions of some energy-intensive manufacturing sectors are positive to petroleum price shocks compared to other sectors that are less energy-intensive. Furthermore, the effects of shocks on large companies were strong and negative in the period preceding the crisis, while medium-sized companies experienced bigger impacts after the financial crisis. Kang et al. (2017) study the relationship between petroleum price shocks and equity returns of the aggregate oil and gas sector index and seven individual companies, namely BP, Chevron, ConocoPhillips, Exxon Mobil, TransCanada Corporation, Royal Dutch Shell, and Valero Energy Corporation, for the period 1985–2015. Their empirical results show that the effects of demand-side shocks on the equity returns of oil and gas companies are significant and positive. In addition, the authors analyse three oil and gas sector components, such as upstream, midstream, and downstream, represented by ConocoPhillips, TransCanada Corporation, and Valero Energy Corporation, respectively. They found that in all three cases, responses to demand-side shocks are positive, whereas petroleum supply shocks cause negative effects. Antonakakis et al. (2018) investigate volatility spillovers between petroleum prices, represented by the WTI sort, and the equity prices of twelve oil and gas companies. The authors employed daily firm-level data from June 2001 to February 2016. Their results demonstrate the significant volatility spillover effects between petroleum and equity prices. In particular, the transmission of volatility is mostly unidirectional, from the equity prices of corporations to petroleum prices. The companies that exerted

the largest impact are British Petroleum, Chevron, Exxon Mobil, Royal Dutch Shell, and Total SA. Furthermore, the authors detect higher dynamic conditional correlations between volatilities during the global financial crisis, while correlation figures are lower during 2005.

10. Intraday Linkages between Petroleum Prices and Stock Markets

Considering the accelerated exchange of information among markets, some studies are devoted to the analysis of the petroleum-stock market nexus utilising intraday data.¹¹ Xu et al. (2019) make use of 5 min data to examine volatility interconnections between the equity markets of China, the United States, and WTI futures prices over the period from 2007 to 2016. Their findings reveal that the volatility connectedness between petroleum and equity markets is stronger for the United States. The dynamic figures of volatility transmissions indicate that the stock market of China and petroleum are the net-recipients throughout the financial crisis, whereas the stock market of the United States is found to be the net-contributor. In addition, the authors observe the prevalence of bad volatility spillovers for most of the study period. Employing a 5 min frequency of data, Suleman et al. (2021) examine asymmetric volatility interactions between the Dow Jones Islamic Market Index and three commodities (gold, silver, and the Brent grade of crude petroleum) for the period from January 2010 to November 2020. The authors document that the Dow Jones Islamic Market Index acts as a major transmitter of volatility for commodities. The dynamic volatility interactions intensify throughout the Eurozone debt crisis and the global health crisis, with Brent crude petroleum and silver being the net-recipients of volatility over the latter event. Furthermore, they found that the intensity of volatility transmissions induced by negative shocks is greater in contrast to positive shocks. Farid et al. (2021) focus on the more recent period from January 2019 to May 2020 to investigate intraday volatility connectedness between equities and commodities employing 5 min trading data on US ETFs. Their results point to the role of the equity market as the major source of volatility transmission to commodities during the entire study period. In addition, the authors observe that the recent pandemic exerted a substantial impact on the total dynamic volatility connectedness, which amplified in the first half of 2020. Heinlein et al. (2021), utilising 5 min data and a contagion test grounded on the local Gaussian correlation approach, report significantly stronger correlations between Brent crude petroleum and equity markets of both petroleum exporting and importing countries during the COVID-19 outbreak. Adekoya et al. (2022) focus on the intraday connectedness between prominent assets, such as Brent crude petroleum, gold, bitcoin, S&P stocks, US bonds, and the dollar index. The authors observe that Brent crude petroleum is the net contributor of spillovers during periods of market instability caused by geopolitical tensions. Applying the GARCH-type model, Mensi et al. (2022) document substantial upsurges in intraday dynamic correlations between returns of the US stock market and Brent crude petroleum throughout the spread of COVID-19.

11. Conclusions

Understanding the interaction mechanisms between petroleum prices and equity markets continues to garner the interest of industry professionals and scholars globally, particularly in view of the unprecedented swings driven by demand-side, supply-side, and speculative factors that petroleum prices experienced over the preceding two decades. By identifying 194 studies published from 1983 to 2023, mostly in top-ranked journals, the current paper delivers a comprehensive review of a broad array of the existing literature by focusing on the following themes: (i) the effects of petroleum prices on macroeconomic variables; (ii) the relationship between petroleum prices and equity returns at the aggregate and sector levels; (iii) the role of petroleum price shocks in elucidating equity returns; (iv) time-varying linkages between petroleum prices and equity markets; (v) volatility spillovers between petroleum prices and equity markets; (vi) the role of a country as a net petroleum exporter or net petroleum importer in the petroleum-equity market nexus; (vii) responses of firm-level equity returns to petroleum price changes; and (viii) intraday

connections between petroleum prices and equity markets. The synthesis and consolidation of an extensive number of empirical works enable researchers and professionals to gain an in-depth understanding of the interaction mechanisms between markets of interest.

The inferences derived from the current survey can be summarised as follows: The effects of petroleum prices and shocks, as well as volatility interactions, are mixed depending on the utilised methodologies, estimation periods, and data frequencies, yet it is evident that, in general, the sensitivities of equity returns are negative for petroleum importers, whereas the opposite trend is valid for petroleum exporters. Negative and positive petroleum price fluctuations are prone to exerting asymmetric effects on equity returns. Furthermore, the nexus between petroleum and equity markets exhibits a dynamic nature and tends to be affected by global events.

Having conducted the comprehensive analysis, it is worth highlighting several avenues that we hope will offer the scope necessary for extending research in this diverse line of literature. Given the growing significance of managing environmental, social, and governance (ESG) risks, it would certainly be interesting to assess the role that sector- and company-specific ESG ratings play when investigating the effects of petroleum price shocks and volatility transmissions. Prior to moving to the next area for extension, it is necessary to walk through the potential effects of petroleum prices on stock markets in petroleum exporting and importing countries. An increase in the price of petroleum induces the relocation of capital to petroleum-exporting nations and generates greater investment opportunities, but for petroleum-importing nations, it results in high manufacturing costs and the amplification of inflationary pressures (Bjornland 2009; Wang et al. 2013; Filis and Chatziantoniou 2014). Thus, not only aggregate markets but also sectors are anticipated to experience different reactions. The utilisation of aggregate market indices in the analysis will not produce a full picture since the level of petroleum dependency is dissimilar among sectors and, hence, the intensity of responses may demonstrate heterogeneous behaviour (Gogineni 2010; Elyasiani et al. 2011). Taking the aforesaid into account, as suggested by Degiannakis et al. (2018), the effects of shocks representing the speculative behaviour in the petroleum market on the petroleum-stock sector nexus across exporters and importers can also be incorporated. However, the existing approach developed by Kilian and Murphy (2014) for detecting speculative shocks based on petroleum inventories data, although useful, is constrained to a monthly frequency. Therefore, it would be beneficial to produce shocks associated with speculative activity by employing high-frequency data, which permits the capture of valuable short-term movements amid the interlinked global financial landscape and, in particular, the enhanced financialization of the petroleum market. Furthermore, another important avenue is associated with quantifying minute-per-minute good and bad volatilities, where a method of decomposing realised variances of assets into realised semivariances proposed by Barndorff-Nielsen et al. (2010) can be applied to scrutinise the intraday volatility interconnectedness at the sector level across petroleum exporters and importers. This will allow us to assess the instant impact of actions taken by OPEC+ countries towards the voluntary reduction of production levels in order to control petroleum prices and the recently increased geopolitical risks, given the accelerated spread of information among financial markets.

Author Contributions: Both authors (M.B. and C.M.) contributed equally to the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Summary of considered studies.

Authors	Countries and Study Periods	Methodology	Variables
Panel A: Petroleum prices and economic variables			
Hamilton (1983)	US 1948–1980	VAR	Petroleum price, real GNP, unemployment, US prices, wages, money, and import prices.
Burbidge and Harrison (1984)	US, Japan, Germany, UK, Canada 1961–1982	VAR	Petroleum price, total industrial production in OECD and domestic countries, short-term interest rate, currency and demand deposits, wages, and CPI.
Gisser and Goodwin (1986)	US 1961–1982	St. Louis-type equations	Petroleum price, money stock, fiscal activity, real GNP, price level, real investment, and unemployment.
Mork (1989)	US 1949–1988	VAR	Petroleum price, real GNP, inflation, 3-month Treasury bill, unemployment, import price, and wages.
Mork et al. (1994)	Canada, France, Germany, Japan, Norway, UK, US 1967–1992	VAR, Seemingly Unrelated Regression	Petroleum price, real GDP, unemployment, wages, and interest rates.
Hooker (1996)	US 1948–1994	VAR	Petroleum prices, 3-month Treasury bill rate, inflation, import prices, unemployment, GDP, industrial production, and employment in goods and services producing sectors.
Hamilton (1996)	US 1948–1994	VAR, Granger causality	Petroleum prices, GDP, Treasury bill, inflation, and import prices.
Bjornland (2000)	US, Germany, UK, Norway 1960–1994	VAR	Petroleum prices, real GDP, and unemployment.
Hamilton (2003)	US 1949–2001	Flexible approach to nonlinear modelling	Petroleum price, GDP.
Cunado and Perez de Gracia (2003)	Germany, Belgium, Austria, Spain, Finland, France, Ireland, Italy, Luxembourg, Portugal, UK, Netherlands, Denmark, Greece, Sweden 1960–1999	VAR	Petroleum price, inflation, and industrial production indices.
Jimenez-Rodriguez and Sanchez (2005)	Canada, France, Germany, Italy, Japan, Norway, UK, US, Euro Area 1972–2001	VAR	Petroleum price, real GDP, exchange rate, wage, inflation, and short and long-term interest rates.
Lardic and Mignon (2006)	Austria, Belgium, Finland, France, Germany, Italy, Netherlands, Norway, Portugal, Spain, Sweden, UK 1970–2003	OLS	Petroleum price, GDP.

Table A1. Cont.

Authors	Countries and Study Periods	Methodology	Variables
Lardic and Mignon (2008)	G7, US, Europe, Euro Area 1970–2004	OLS	Petroleum price, GDP.
Hou et al. (2016)	Canada 1980–2011	SVAR	Petroleum price, GDP, inflation, wage, and exchange rate.
Lorusso and Pieroni (2018)	UK 1976–2014	SVAR	Brent, real GDP, output gap, inflation, and short-term interest rate.
Wen et al. (2021)	G7 1997–2019	Diebold and Yilmaz	NYMEX Crude-Light Sweet Oil, VIX, the world integrated oil and gas producer index, and CPI.
Cunado and Perez de Gracia (2005)	Japan, Singapore, South Korea, Malaysia, Thailand, Philippines 1975–2002	Cointegration, Granger causality, VAR	Petroleum price, CPI, Industrial Production Index (Japan, South Korea), Manufacturing Production Index (Singapore), and real GDP (Malaysia, Thailand, and Philippines).
Kim et al. (2017)	China 1992–2014	TVP SVAR, SVAR, GIR VAR	Petroleum price, petroleum production, industrial production, CPI, exchange rate, and interest rate.
Liu et al. (2020)	China 1999–2018	SVAR	WTI, industrial added value growth rate, money supply, and CPI.
Lin et al. (2023)	Euro-19, China, Japan, US 1990–2021	BVARSV	WTI, Brent, exchange rate, interest rate, consumer price index, industrial production index (Euro-19, Japan), industrial added value (China), and gross industrial output (US).
Korhonen and Ledyaeva (2010)	Russia, Germany, Italy, Netherlands, China, US, UK, Switzerland, Finland, Belgium, Canada, France, Japan 1995–2006	VAR	WTI, Brant, Dubai, and GDP.
Nusair (2016)	Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, UAE 1968–2014	NARDL	Dubai, real GDP.
Lee et al. (2017)	Canada, UK, Germany, France, Italy, Japan, US 1994–2014	SVAR	World petroleum production, real global economic activity, Brent, and risk ratings.
Smiech et al. (2021)	Canada, Mexico, Norway, Russia 2006–2019	BSVAR	WTI, industrial production, and exchange rate.
Tan and Uprasen (2023)	Brunei, Indonesia, Malaysia, Vietnam, Myanmar, Philippines, Singapore, Thailand 2000–2021	Panel NARDL	Petroleum price, Gini index, CPI, GDP, unemployment, human development, and trade openness.

Table A1. Cont.

Authors	Countries and Study Periods	Methodology	Variables
Panel B: Petroleum prices and aggregate stock markets			
Jones and Kaul (1996)	US, Canada, Japan, UK 1947–1991	Standard cash-flow dividend valuation model	Petroleum price, index of industrial production, inflation, equity returns, dividend yield, corporate bond yield, government bond yield, default spread, and term spread.
Sadorsky (1999)	US 1947–1996	VAR	Petroleum price, S&P 500, industrial production, and interest rate.
Papapetrou (2001)	Greece 1989–1999	VAR	Petroleum price, equity returns, interest rate, industrial production, and employment.
Driesprong et al. (2008)	48 countries 1973–2003	Basic regression model	Brent, WTI, Dubai, Arab Light, IPE Brent futures, NYMEX Light futures, and MSCI reinvestment indices.
Chen (2010)	US 1957–2009	TVTP-MS	Petroleum price, S&P500.
Filis (2010)	Greece 1996–2008	VECM, VAR	Brent, ATHEX General Composite Index, industrial production, and CPI.
Basher et al. (2012)	Emerging markets 1988–2008	SVAR	Petroleum price, petroleum production, global real economic activity, MSCI emerging stock market index. exchange rate, and interest rate.
Asteriou and Bashmakova (2013)	Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Russia, Slovakia, Slovenia 1997–2007	International multi-factor model	WTI, stock markets, exchange rate, and MSCI World Index.
Narayan and Narayan (2010)	Vietnam 2000–2008	OLS, DOLS, GARCH	WTI, stock market, and exchange rate.
Zhu et al. (2011)	14 OECD and Non-OECD markets 1995–2009	Panel threshold cointegration, TVECM, Granger causality	WTI, stock prices, industrial production, and short-term interest rate.
Zhu et al. (2014)	Australia, China, Hong Kong, India, Indonesia, Japan, South Korea, Malaysia, Singapore, Taiwan 2000–2012	AR-GARCH, Constant and time-varying copulas	WTI, S&P / ASX 200, Shanghai composite, Hang Seng, BSE National, Jakarta SE Composite, Nikkei 225, Kospi, Kuala Lumpur Composite, Strait Times, and SE Weighted.
Silvapulle et al. (2017)	China, France, Germany, India, Italy, Japan, Singapore, South Korea, Spain, US 1995–2015	Nonparametric panel data model	WTI, S&P500, Nikkei 225, Shanghai SE Composite, KOSPI 200, S&P BSE 30, DAX 30, CAC 40, Straits Times, FTSE MIB, and IBEX 35.
Huang et al. (1996)	US 1979–1990	VAR	NYMEX petroleum futures, S&P 500.
Basher and Sadorsky (2006)	21 emerging markets 1992–2005	OLS, Unconditional and conditional cross section regressions	WTI, MSCI World Index, and exchange rate.

Table A1. Cont.

Authors	Countries and Study Periods	Methodology	Variables
Cong et al. (2008)	China 1996–2007	VAR	Brent, industrial production, CPI, short term interest rate, and stock market prices.
O'Neill et al. (2008)	Australia, Canada, France, UK, US 2003–2006	ARX	WTI, DJ Industrial Average, TSX, FTSE 100, and CAC 40.
Miller and Ratti (2009)	France, Germany, Italy, UK, Canada, US 1971–2008	VECM	Brent, CPI, producer price index, CPI, industrial production index, short-term interest rate, and stock market prices.
Jammazi and Aloui (2010)	France, Japan, UK 1989–2007	MS-VAR	WTI, Brent, FTSE 100, CAC 40, and Nikkei 225.
Ajmi et al. (2014)	Bahrain, Egypt, Jordan, Kuwait, Lebanon, Morocco, Oman, Qatar, Saudi Arabia, Tunisia, UAE 2007–2012	Granger causality, Mackey-Glass process, and Kyrtsou-Labys causality	WTI, Brent, and stock market prices.
Narayan and Gupta (2015)	US 1859–2013	Predictive regression model	WTI, S&P 500.
Hatemi et al. (2017)	US, UK, France, Germany, Italy, Canada, Japan, World 1975–2013	Granger causality, Asymmetric causality of Hatemi-J	Petroleum prices, stock market prices.
Panel C: Petroleum prices and stock sectors			
Faff and Brailsford (1999)	Australia 1983–1996	Two-factor APT	Petroleum prices, 24 ASX sectors.
Sadorsky (2001)	Canada 1983–1999	Multifactor market model	WTI, TSE oil and gas index.
Boyer and Filion (2007)	Canada 1995–2002	GLS	WTI, NYMEX Natural Gas, stock market, interest rate, exchange rate, cash flows, debt, petroleum production, proven reserves, and drilling success.
Hammoudeh and Li (2005)	US 1986–2003	VECM	WTI, NYSE Transportation Index, Amex Oil Index, and MSCPI Index.
El-Sharif et al. (2005)	United Kingdom 1989–2001	Two-factor model	Brent, oil and gas sector index, exchange rate, FTSE All Share Index, and interest rate.
Lee et al. (2012)	Canada, France, Germany, Italy Japan, UK, US 1991–2009	VAR	Petroleum price, industrial production, interest rate, Composite, Consumer discretionary, Consumer staples, Energy, Financial, Health care, Industrials, Information technology, Materials, Utilities, Transportation, and Telecommunications sector indices.

Table A1. Cont.

Authors	Countries and Study Periods	Methodology	Variables
Moya-Martinez et al. (2014)	Spain 1993–2010	Multifactor market model	Brent, aggregate market index, interest rate, Banking, Basic Resources, Chemicals and Paper, Construction, Consumer Goods, Consumer Services, Energy, Financial, Food and Beverages, Health Care, Industrials, Real Estate, Services, Technology and Telecommunications, and Utilities sectors.
Xu (2015)	UK 1988–2013	One- and two-factor models, Diebold-Mariano	Brent, Basic Materials, Consumer Goods, Consumer Services, Financials, Health Care, Industrials, Oil and Gas, Technology, and Telecom and Utilities FTSE All-Share industry indices.
Arouri (2011)	Europe 1998–2010	Linear and asymmetric models	Brent, DJ Stoxx 600, Automobile and Parts, Financials, Food and Beverages, Oil and Gas, Health Care, Industrials, Basic Materials, Personal and Household Goods, Consumer Services, Technology, Telecommunications, and Utilities DJ Stoxx sector indices.
Scholtens and Yurtsever (2012)	Euro Area 1983–2007	VAR, MLRM	Brent, short-term interest rate, industrial production, and 38 industries.
Li et al. (2012)	China 2001–2010	Panel cointegration, Granger causality	WTI, Agriculture, Conglomerates, Construction, Financials, IT, Manufacturing, Media, Mining, Real Estate, Social services, Transportation, Utilities, and Wholesale and Retail sectors.
Zhu et al. (2016)	China 1994–2014	Quantile regression approach	WTI, 16 sectors (Accommodation and Catering; Agriculture; Complex; Construction; Culture, Sports and Entertainment; Financial; IT; Manufacturing; Mining; Production and Supply of Power, Heat, Gas and Water; Realty; Transportation; Water, environment and public facilities management; Wholesale and retail trade).
Broadstock et al. (2014)	Japan, India, Korea, Taiwan 1984–2012	Simple empirical asset pricing model	WTI, KOSPI, KOSPI 200 Energy and Chemical, NKY, SENSEX, SENSEX Oil and Coal, SENSEX Power, TOPIX, TOPIX Oil, Nikkei 225, Nikkei 500 Oil, TWSE, and Taiwan TaieX Oil.
Nandha and Faff (2008)	Global 1983–2005	Standard market model	WTI, 35 DataStream Global Industry indices

Table A1. Cont.

Authors	Countries and Study Periods	Methodology	Variables
Ramos and Veiga (2011)	34 countries 1998–2009	Multifactor panel model	Brent, world market index, oil and gas industry indices, currency rates, and interest rates.
Henriques and Sadorsky (2008)	US 2001–2007	VAR	WTI, WilderHill Clean Energy Index, and Arca Tech 100 index.
Kumar et al. (2012)	Global 2005–2008	VAR	WTI, Brent, Wilder Hill New Energy Global Innovation Index, Wilder Hill Clean Energy Index, S&P Global Clean Energy Index, S&P 500 Index, carbon price, and interest rate.
Cameron and Schnusenberg (2009)	Global 2001–2008	Four-factor regression model	WTI, Energy Select Sector SPDR (XLE) ETF, stock prices of General Motors, Ford Motor Corp., Daimler/Chrysler, Toyota Motor Corp., Honda Motor Co., and Nissan Motors.
Nandha and Brooks (2009)	38 countries 1983–2006	Standard market model	WTI, transport sector indices, and world market index.
Kristjanpoller and Concha (2016)	Global 2008–2013	CAPM, GARCH	WTI, Jet Fuel, and 56 airlines’ stock prices.
Panel D: Petroleum price shocks and stock markets			
Kilian and Park (2009)	US 1973–2006	SVAR	US aggregate and 4 sector (Petroleum and Natural Gas, Automobiles and Trucks, Retail, Precious Metals) stock returns, oil supply shock, aggregate demand shock, and oil-specific demand shock.
Kang et al. (2016)	1973–2014	SVAR	US aggregate and 4 sector (Petroleum & Natural Gas, Automobiles and Trucks, Retail, Precious Metals) stock returns, US oil supply shock, non-US oil supply shock, aggregate demand shock, and oil-specific demand shock.
Apergis and Miller (2009)	Canada, US, France, UK, Germany, Italy, Australia, Japan 1981–2007	VEC	Australian General Market Index, C.L. Toronto Index, DAX Index, CAC Industrial Index, MIB 30 Index, Nikkei Index, FT 30 Index, NYSE Index, oil supply shock, global aggregate demand shock, and global oil demand.
Kang and Ratti (2013)	US 1985–2011	SVAR	US aggregate and 4 sector stock returns, US economic policy uncertainty, oil supply shock, aggregate demand shock, and oil market specific demand shock.
Abhyankar et al. (2013)	Japan 1988–2009	SVAR	Datastream Japan country equity index, oil supply shock, aggregate demand shock, and oil market specific demand shock.

Table A1. *Cont.*

Authors	Countries and Study Periods	Methodology	Variables
Guntner (2014)	Canada, France, Germany, Japan, Norway, US 1974–2011	SVAR	S&P 500, CDAX, NIKKEI 225, S&P/TSX, CAC40, OBX, oil supply shock, aggregate demand shock, and other oil demand shock.
Angelidis et al. (2015)	US 1989–2011	SVAR	Dow Jones returns, Dow Jones realised volatility, dividend yield, interest rate, unemployment, default spread, inflation, petroleum production, petroleum price returns, global economic activity, and petroleum price volatility.
Bastianin et al. (2016)	Canada, US, France, UK, Germany, Italy, Japan 1973–2015	SVAR	MSCI country indices, oil supply shock, aggregate demand shock, and oil specific demand shock.
Sakaki (2019)	US 1990–2015	SVAR	10 S&P 500 sector indices (Consumer Discretionary, Consumer Staples, Energy, Financials, Health Care, Industrials, Information Technology, Materials, Telecommunication Services, Utilities), oil supply shock, aggregate demand shock, and oil market specific demand shock.
Hwang and Kim (2021)	US 1973–2008	Smooth Transition VAR	S&P 500, 4 industries (Petroleum and Natural Gas, Chemicals, Automobile and Truck, Retail) oil supply shock, aggregate demand shock, and oil specific demand shock.
Gupta and Modise (2013)	South Africa 1973–2011	SVAR	Johannesburg Securities Exchange All Share Index, oil supply shock, global demand shock, and speculative demand shock.
Fang and You (2014)	China, India, Russia 2001–2012	SVAR	Aggregate market indices, oil supply shock, global demand shock, and oil specific demand shocks.
Li et al. (2017)	China 1994–2014	SVAR	Listed firms from the petroleum industrial chain in China, oil supply shock, global demand shock, precautionary demand shock, and domestic demand shock.
Koh (2017)	Bangladesh, China, Hong Kong, India, Indonesia, Japan, Korea, Malaysia, Nepal, Pakistan, Philippines, Singapore, Sri Lanka, Taiwan, Thailand 1994–2014	SVAR	Stock markets, oil supply shock, aggregate demand shock, and oil market specific demand shock.
Demirer et al. (2020)	21 countries 2000–2018	Multifactor linear model, Diebold and Yilmaz	Stock markets, MSCI world stock index, FTSE world government bond index, oil demand shock, oil supply shock, and risk shock.

Table A1. Cont.

Authors	Countries and Study Periods	Methodology	Variables
Wong (2020)	China 2000–2019	Multifactor model, Three-factor model	Forms from 11 sectors and 60 industries, oil demand shock, oil supply shock, and risk shock.
Mishra and Mishra (2021)	India 2010–2019	Linear regression, Rolling window regression, DCC GARCH, and MRS	10 NSE sector indices (Auto, Bank, Consumer Durables, FMCG, Financial Services, Information Technology, Media, Metal, Pharma and Realty), oil demand shock, oil supply shock, and risk shock.
Umar et al. (2020)	Spain 2000–2019	Diebold and Yilmaz	8 sectors (Basic Materials, Industrials, Consumer Goods, Telecommunications, Utilities, Financials, Technology and Retail), oil demand shock, oil supply shock, and risk shock.
Umar et al. (2021)	Bahrain, Brazil, China, India, Kuwait, Oman, Qatar, Russia, South Africa, Saudi Arabia, UAE 2005–2020	Diebold and Yilmaz	MSCI total return stock indices, oil demand shock, oil supply shock, and risk shock.
Panel E: Time-varying linkages between petroleum prices and stock markets			
Reboredo (2010)	Germany, Netherlands, UK, US 1985–2006	MS models	Brent, WTI, Dubai, S&P 500, DAX, FTSE 100, and AEX.
Mollick and Assefa (2013)	US 1999–2011	GARCH, MGARCH-DCC	WTI, gold, exchange rate, VIX, inflation, interest rates, Russell 2000, NASDAQ, Dow Jones, and S&P 500.
Chang and Yu (2013)	US 2001–2012	MS-ARJI-GJR-GARCH-X	WTI, S&P 500.
Ciner (2013)	US 1986–2010	Frequency domain method	WTI, NASDAQ, S&P 500, and 20 stocks from the DJIA.
Degiannakis et al. (2013)	Europe 1992–2010	Diag-VECH GARCH	Brent, 10 Dow Jones sector indices (Basic Materials, Consumption Goods, Financials, Health, Industrial, Oil and Gas, Retail, Technology, Telecommunications, and Utilities).
Reboredo and Rivera-Castro (2014)	2000–2011	Wavelet multi-resolution analysis	Brent, S&P 500, DJ Stoxx Europe 600, 8 sector indices (Automobile and Parts, Banks, Chemical, Oil and Gas, Industrial Good, Utilities, Telecommunications, and Technologies).
Martin-Barragan et al. (2015)	US, Germany, UK, Japan 1990–2011	Wavelet-based approach	WTI, aggregate market indices.
Bhar and Nikolova (2010)	Russia 1995–2007	EGARCH	WTI, AK&M Composite index
Mohanty et al. (2010)	Hungary, Czech Republic, Romania, Poland, Slovenia, Austria 1998–2010	Two-factor model	WTI, major oil and gas companies, and aggregate market.

Table A1. Cont.

Authors	Countries and Study Periods	Methodology	Variables
Broadstock et al. (2012)	China 2000–2011	BEKK-GARCH, Three-factor model	Brent, energy sector index.
Broadstock and Filis (2014)	China, US 1995–2013	SVAR, Scalar-BEKK	NYSE, Shanghai Composite, 5 sector indices (Banks, Retail, Oil and Gas, Metals and Mining, Technology), oil supply shock, aggregate demand shock, and oil market specific demand shock.
Reboredo and Ugolini (2016)	US, UK, European Monetary Union, Brazil, Russia, India, China, South Africa 2000–2014	Unconditional and conditional quantiles, TGARCH, and Copula models	Brent, aggregate market indices.
Zhang (2017)	China, US 2002–2014	MADCC	WTI, CSI 300, and S&P 500.
Zhu et al. (2017)	China, India, Japan, South Korea, US, Canada, Mexico, Russia, UK 1997–2015	SVAR, Two-stage Markov regime-switching	Shanghai Composite, BSE 30, Nikkei 225, Seoul Composite, S&P 500, BOVESPA, S&P TSX Composite, MX-MEXBOL, RTS, FTSE 100, oil supply shock, aggregate demand shock, and oil demand shock.
Panel F: Petroleum price and stock market volatility			
Aloui and Jammazi (2009)	France, UK, Japan 1989–2007	Two-regime MS-EGARCH	WTI, Brent, CAC 40, FTSE 100, and Nikkei 225.
Vo (2011)	US 1999–2008	CC-MSV, DC-MSV	WTI futures, S&P 500.
Mensi et al. (2013)	US 2000–2011	VAR-GARCH	WTI, Brent, S&P 500, gold, wheat, and beverage price indices.
Chang et al. (2013)	UK, US 1998–2009	Multivariate GARCH (CCC, VARMA-GARCH, VARMA-AGARCH, and DCC)	Brent and WTI spot prices, Brent and WTI one-month forward prices, Brent one-month futures prices, NYMEX one-month futures prices. S&P 500, NYSE, Dow Jones, and FTSE 100.
Kang et al. (2015)	US 1973–2013	SVAR	AMEX, NYSE, Nasdaq, VIX, oil supply shock, aggregate demand shock, and oil market specific demand shock.
Salisu and Oloko (2015)	US 2002–2014	VARMA-BEKK-AGARCH	WTI, Brent, and S&P 500.
Khalifaoui et al. (2015)	Canada, France, Germany, Japan, UK, US 2003–2012	Wavelet-based MGARCH	WTI, S&P/TSX, CAC 40, DAX, FTSE MIB, Nikkei 225, FTSE 100, and S&P 500.
Ewing and Malik (2016)	US 1996–2013	Univariate GARCH, BEKK-GARCH	WTI, S&P 500.
Adekoya and Oliyide (2021)	US January 2020–July 2020	TVP-VAR	Crude petroleum S&P 500, bitcoin, gold, and exchange rate.
Malik and Hammoudeh (2007)	US, Bahrain, Kuwait and Saudi Arabia 1994–2001	BEKK-GARCH	WTI, S&P 500, BSE, KSE, and Tadawul.

Table A1. Cont.

Authors	Countries and Study Periods	Methodology	Variables
Maghyereh et al. (2016)	Canada, Germany, India, Japan, Mexico, Russia, South Africa, Sweden, Switzerland, UK, US 2008–2015	Diebold and Yilmaz	OVX, AEXVOLI, VIXCVOLI, VFTSEIX, NIFVIXI, VIMEXVI, VXJINDX, SIXVXVL, RTSVXVL, JSAVIVI, VDAXNEW, and VSMI01M.
Tiwari et al. (2021)	Developed and emerging markets 2000–2017	Diebold and Yilmaz	Brent, WTI, NYK, DAX, SPX, UKK, CAC40, FTSEMIB, SPTSX, MICEX, IBOV, SENSEX, Shanghai, and TOP40.
Mensi et al. (2021)	Brazil, Russia, India, China, South Africa, US, Japan, Australia, Canada, France, Germany 2000–2018	Baruník and Krehlík connectedness approach, Wavelet-based approach, and DCC-GARCH	WTI futures, gold futures, IBOV, RTS, SENSEX, SHCOMP, JALSH, SPX, Nikkei, All Ords, TSX, FTSE, CAC, and DAX.
Arouri et al. (2011a)	Bahrain, Kuwait, Oman, Saudi Arabia, Qatar, UAE 2005–2010	VAR-GARCH	Brent, stock market indices.
Awartani and Maghyereh (2013)	Bahrain, Kuwait, Oman, Saudi Arabia, Qatar, UAE 2004–2012	Diebold and Yilmaz	WTI, stock market indices
Lin et al. (2014)	Nigeria, Ghana 2000–2010	DCC-GARCH, VAR-GARCH, and VAR-AGARCH	Brent, GSE All-Share Index, and NSE All-Share Index.
Yousaf and Hassan (2019)	India, China, Indonesia, Korea, Malaysia, Pakistan, Philippines, Taiwan, Thailand 2000–2018	VAR-GARCH, VAR-AGARCH	Petroleum price, aggregate market indices.
Al-Yahyaee et al. (2019)	Bahrain, Kuwait, Oman, Saudi Arabia, Qatar, UAE 2005–2016	DECO-FIAPARCH, Diebold and Yilmaz	WTI, gasoline, heating oil, gold, palladium, platinum, silver, and aggregate market indices.
Sarwar et al. (2020)	China, India, Pakistan 1997–2015	BEKK-GARCH	WTI, Shanghai, and Karachi and Bombay stock markets.
Hammoudeh et al. (2004)	US 1995–2001	VAR, VECM, univariate GARCH, and multivariate GARCH	WTI, S&P Oil Composite index, S&P Oil Domestic Integrated index, Oil and Gas Exploration index, S&P Oil and Gas (Refining and Marketing) index, and S&P Oil-International Integrated index.
Malik and Ewing (2009)	US 1992–2008	BEKK-GARCH	WTI, DJ sector indices (financials, technology, consumer services, health care, and industrials).
Elyasiani et al. (2011)	US 1998–2006	AR-GARCH	NYMEX petroleum futures, 13 US sectors.
Arouri et al. (2011b)	US, Europe 1998–2009	CCC-GARCH, BEKK-GARCH, DCC-GARCH, and VAR-GARCH	Brent, S&P 500, DJ Stoxx Europe 600, and 7 sector indices (Automobile and Parts, Basic Materials, Financials, Industrials, Telecommunications, Technology, and Utilities).

Table A1. Cont.

Authors	Countries and Study Periods	Methodology	Variables
Arouri et al. (2012)	Europe 1998–2009	VAR-GARCH	Brent, DJ Stoxx Europe 600, 7 sector indices (Automobile and Parts, Basic Materials, Financials, Industrials, Telecommunications, Technology, and Utilities).
Sadorsky (2012)	US 2001–2010	Multivariate GARCH (BEKK, diagonal, CCC, and DCC)	WTI, WilderHill Clean Energy Index, and Arca Technology Index.
Degiannakis et al. (2014)	Europe 1999–2010	SVAR	Euro Stoxx 50, 10 DJ sector indices (Basic Materials, Consumption Goods, Financials, Health, Industrial, Oil and Gas, Retail, Technology, Telecommunications, Utilities), supply-side oil shock, aggregate demand shock, oil specific demand shock, and volatility shock.
Als Salman (2016)	US 1973–2014	Bivariate GARCH-in-mean VAR	Petroleum price, aggregate market index, and 18 sectors.
Belhassine (2020)	Europe 2004–2015	VAR-BEKK-GARCH	Brent, Euro STOXX 50, and 19 Eurozone supersector indices.
Costola and Lorusso (2022)	Russia 2005–2020	Diebold and Yilmaz	Brent, natural gas, gold, MSCI aggregate market indices, MOEX, six sector indices (Telecommunications, Financials, Oil and Gas, Metals and Mining, Electric Utilities, and Consumers Goods and Services).
Panel G: Petroleum prices and stock markets in petroleum exporting and importing countries			
Park and Ratti (2008)	US, Austria, Belgium, Denmark, Finland, France, Germany, Greece, Italy, Netherlands, Norway, Spain, Sweden, UK (importers) 1986–2005	VAR	Petroleum price, stock markets, interest rate, industrial production, and CPI.
Ramos and Veiga (2013)	Austria, Belgium, Finland, France, Germany, Greece, Ireland, Japan, the Netherlands, Portugal, Spain, Sweden, Switzerland (importers); Colombia, Canada, Mexico, Norway, Russia (exporters) 1988–2009	Fixed effects, GARCH	NYMEX petroleum futures, stock markets, world market index, and exchange rate.
Wang et al. (2013)	China, France, Germany, India, Italy, Japan, Korea, UK, US (importers); Canada, Saudi Arabia, Kuwait, Mexico, Norway, Russia, Venezuela (exporters) 1999–2011	SVAR	S&P 500, NIKKEI 225, DAX, CAC 40, FTSE 100, FTSE MIB, Shanghai Composite, KOSPI Composite, BSE Sensex, Tadawul All Share, Kuwait Stock Exchange Index, Bolsa IPC, OSEAX, MICEX, IBVC, S&P/TSX Composite, oil supply shock, aggregate demand shock, oil specific shock, and stock specific shock.

Table A1. *Cont.*

Authors	Countries and Study Periods	Methodology	Variables
Le and Chang (2015)	Malaysia (exporter), Japan (importer), Singapore (refinery) 1997–2013	VAR, Toda and Yamamoto causality approach	Dubai crude petroleum, stock markets, interest rate, and industrial production.
Wang and Liu (2016)	China, France, Germany, India, Japan, Korea, UK, US, Spain (importers); Canada, Saudi Arabia, Kuwait, Mexico, Norway, Russia, Venezuela (exporters) 2000–2011	BEKK-GARCH, CCC-GARCH, DCC-GARCH, and RS-DCC	WTI, SSEC, FCHI, GDAXI, BSESX, NIKKEI 225, KS11, FTSE, S&P 500, TSX, TASI, SEWI, MXX, OSEAX, MICEX, SMSI, and IBVC.
Salisu and Isah (2017)	Argentina, Australia, France, Germany, Japan, South Korea, UK, US (importers); Kuwait, Indonesia, Nigeria, Qatar, Saudi Arabia (exporters) 2000–2015	Nonlinear Panel ARDL	Brent, WTI, and stock markets.
Ashfaq et al. (2019)	China, Japan, India, South Korea (importers); Saudi Arabia, United Arab Emirates, Iraq (exporters) 2009–2018	BEKK-GARCH, DCC-GARCH, ABEKK-GARCH, and ADCC-GARCH	Petroleum price, stock exchanges.
Belhassine and Karamti (2021)	China, India, US (importers); Canada, Russia, Saudi Arabia (exporters) 2001–2017	Wavelet-based multivariate GARCH	Petroleum price, SZSE Component index, BSE Sensex index, S&P 500 index, TSXSP, RTSI, and TASI.
Enwereuzoh et al. (2021)	Botswana, Kenya, Mauritius, South Africa (importers); Egypt, Tunisia, Nigeria (exporters) 2000–2018	SVAR, Two-state regime smooth transition regression framework	Botswana Gaborone Index, Nairobi Securities Exchange Index, Mauritius Stock Exchange Index, Johannesburg All Share Index, Egyptian Exchange Index, Tunisian Stock Exchange Index, Nigerian Stock Exchange Index, oil supply shock, aggregate demand shock, and oil specific demand shock.
Akyildirim et al. (2022)	29 exporters and importers 2006–2021	TVP-VAR, Quantile regression approach	MSCI energy stock indices, Twitter-based economic uncertainty index, News Sentiment Index, Infectious Disease Equity Market Volatility Tracker, CBOE Energy Sector ETF Volatility Index, and lockdown index indicator.
Filis et al. (2011)	Netherlands, Germany, US (importers); Brazil, Mexico, Canada (exporters) 1987–2009	DCC-GARCH-GJR	Brent, AEX General Index, DAX 30, Dow Jones Industrial, Bovespa Index, MXICP 35, and S&P/TSX 60.
Antonakakis and Filis (2013)	Germany, UK, US (importers); Norway, Canada (exporters) 1988–2011	DCC-GARCH	Brent, DAX 30, FTSE 100, Dow Jones, OBX, and TSX.

Table A1. Cont.

Authors	Countries and Study Periods	Methodology	Variables
Guesmi and Fattoum (2014)	France, Netherlands, Germany, Italy, US (importers); Venezuela, Saudi Arabia, Kuwait, UAE (exporters) 2000–2010	GJR-DCC-GARCH model	Brent, stock market indices.
Boldanov et al. (2016)	China, Japan, US (importers); Canada, Norway, Russia (exporters) 2000–2014	Diag-BEKK	Brent, TSX, RTS, OSEAX, S&P 500, SSE, and Nikkei 225.
Mokni (2020)	China, Japan, India, South Korea (importers); Russia, Norway, Venezuela, Mexico (exporters) 1999–2018	SVAR, OLS, and TVP models	SSE, NIKKEI225, BSE, KSE, RTSI, OBX, AAPL, IPC, oil supply shock, aggregate demand shock, and oil specific demand shock.
Masih et al. (2011)	South Korea (importer) 1988–2005	VEC	Petroleum price, Korean stock market index, industrial production, and interest rate.
Cunado and Perez de Gracia (2014)	Austria, Belgium, Denmark, Finland, France, Germany, Italy, Luxembourg, Netherlands, Spain, Portugal, UK (importers) 1973–2011	VAR, VECM	Brent, stock markets, industrial production, and interest rate.
Bouri (2015)	Jordan, Lebanon, Morocco, Tunisia (importers) 2003–2013	ARMAX–GARCH	Brent, stock market indices, and MSCI world index.
Silvapulle et al. (2017)	China, France, Germany, India, Italy, Japan, Singapore, South Korea, Spain, US (importers) 1999–2015	Fixed effects with time-varying trend, Parametric panel data model	WTI, S&P500, Nikkei 225, Shanghai SE, KOSPI 200, S&P BSE 30, DAX 30, CAC 40, Strait Times, FTSE MIB, IBEX 35, unemployment rate, bond rate, market capitalisation, dividend yield, and price earnings ratio.
Sarwar et al. (2019)	China, Japan, India (importers) 2000–2016	GO-GARCH, DCC-GARCH, cDCC-GARCH, and BEKK-GARCH	WTI, Bombay stock exchange, Nikkei stock exchange, and Shanghai stock exchange.
Bjornland (2009)	Norway (exporter) 1993–2005	SVAR	OSEBX, unemployment, interest rate, exchange rate, and oil price shocks.
Mohanty et al. (2011)	Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, UAE (exporters) 2005–2009	Two-factor model	WTI, world stock market index, and aggregate stock markets and sectors.
Arouri and Rault (2012)	Bahrain, Kuwait, Oman, Saudi Arabia (exporters) 1996–2007	Bootstrap panel cointegration, seemingly unrelated regression	OPEC spot prices, stock markets.
Demirer et al. (2015)	Kuwait, Qatar, UAE, Saudi Arabia (exporters) 2004–2013	Multifactor model	Brent, stock exchanges.
Gil-Alana and Yaya (2014)	Nigeria (exporter) 2000–2011	Fractionally cointegrated framework	Petroleum prices, Nigerian All Share Index.

Table A1. Cont.

Authors	Countries and Study Periods	Methodology	Variables
Basher et al. (2018)	Canada, Mexico, UK, Norway, Russia, Saudi Arabia, Kuwait UAE (exporters) 1974–2015	SVAR, Multifactor model, and Multifactor Markov-switching model	Stock market indices, oil supply shock, global demand shock, oil specific demand shock, and oil inventory shock.
Bouri et al. (2016)	Jordan (importer) 2004–2013	ARMAX-EGARCH, ARMAX-GARCH, and VARMA-BEKK-AGARCH	Brent, Amman stock exchange, and 3 sector indices (Industrials, Financials, and Services).
Bagirov and Mateus (2022)	Mexico (exporter); UK (importer) 2005–2018	VAR-GARCH	Brent, S&P BMV IPC, FTSE 100, and 5 sector indices (Basic Materials, Consumer Cyclicals, Consumer Non-Cyclical, Financials, and Industrials).
Panel H: Petroleum prices and firm-level stock returns			
Sadorsky (2008)	US 1990–2006	Multifactor market model	WTI (price and volatility), S&P 1500 index and stocks, size, and spread.
Narayan and Sharma (2011)	US 2000–2008	GARCH, Multiple threshold regression	Petroleum, 560 firms from 14 sectors listed on the NYSE, exchange rate, and turnover rate.
Dayanandan and Donker (2011)	US 1990–2008	GMM	WTI, 200 large oil and gas firms listed on the US Stock Exchange, ROE, size, and gearing.
Aggarwal et al. (2012)	US 1986–2008	Cross-sectional regression, Two-factor model	WTI, constituents of the S&P Transportation industry index, profitability, investment growth, leverage, size, runup, and industry concentration.
Mohanty et al. (2013)	US 1986–2008	Two-factor model	WTI, 54 companies from the US oil and gas sector, size, financial leverage, profitability, growth potential, runup, and industry concentration,
Phan et al. (2015)	US 1986–2010	GARCH	WTI, top 20 companies from air transport, chemical manufacturing, construction, chemical manufacturing, petroleum sub-sectors, top 60 companies in CONGEP sub-sector, and aggregate stock market.
Gupta (2016)	Global 1983–2014		NYMEX petroleum futures, 2136 stocks from 70 countries, Market Dislocation Index, aggregate market, firm riskiness, illiquidity, market value, price to book value, leverage, risk-free rate, foreign sales percentage, and GPD growth rate.
Bagirov and Mateus (2019)	Western Europe 2005–2014	GMM, Panel least squares, Fixed effects, and Random effects	Brent, 137 listed firms and 531 unlisted oil and gas firms, size, and gearing profitability.
Narayan and Sharma (2014)	US 2000–2008	GARCH	Crude petroleum price growth rate, 560 firms from 14 sectors listed on the NYSE, and size.

Table A1. Cont.

Authors	Countries and Study Periods	Methodology	Variables
Tsai (2015)	US 1990–2012	OLS with panel-corrected standard errors	WTI, 682 firms, exchange rate, gold, federal funds rate, yield rate, aggregate stock market, and size.
Kang et al. (2017)	Global 1985–2015	SVAR	Aggregate oil and gas sector index, BP, Chevron, ConocoPhillips, Exxon Mobil, TransCanada Corporation, Royal Dutch Shell, Valero Energy Corporation, policy uncertainty index, oil supply shock, aggregate demand shock, and oil-specific demand shock.
Antonakakis et al. (2018)	Global 2001–2016	DCC GARCH, Diebold and Yilmaz	WTI, BP, Chevron, CNCP, Eni, Exxon Mobil, Lukoil, Petrobras, Royal Dutch Shell, Sinopec, Statoil, Total, and Valero Energy Corporation.
Panel I: Intraday linkages between petroleum prices and stock markets			
Xu et al. (2019)	China, US 2007–2016, 5 min data	Diebold and Yilmaz	WTI futures, Shanghai Stock Exchange Composite Index, and S&P 500 Index.
Suleman et al. (2021)	Middle East 2010–2020, 5 min data	Diebold and Yilmaz	Brent, gold, silver, and Dow Jones Islamic Market Index.
Farid et al. (2021)	US January 2019–May 2020, 5 min data	MCS GARCH, Diebold and Yilmaz	SPDR S&P 500 trust ETF, US oil fund ETF, US natural gas fund ETF, SPDR gold shares trust ETF, and iShares silver trust ETF.
Heinlein et al. (2021)	Japan, China, Sweden, Canada, Russia, Norway August 2019–April 2020, 5 min data	Local Gaussian correlation	Brent, SPTSX60, NKY, IMOEX, SHCOMP, OSEAX, and OMX.
Adekoya et al. (2022)	US January 2022–March 2022, 30 min data	TVP-VAR	Brent, gold, bitcoin, US bonds, US dollar index, and S&P stocks.
Mensi et al. (2022)	US April 2018–April 2020, 15 min data	FIAPARCH-DCC	Brent futures, gold futures, and S&P 500.

- Notes**
- ¹ British Petroleum (2022).
 - ² We acknowledge that the applied approach could potentially be improved by giving more emphasis to recent papers published in the last 5 years, given unprecedented petroleum price swings, to broadly reflect the current state of knowledge.
 - ³ Panel A of Table A1 in Appendix A summarises the studies analysed in this section.
 - ⁴ Panel B of Table A1 in Appendix A summarises the papers reviewed in this section.
 - ⁵ Panel C of Table A1 in Appendix A summarises the works scrutinised in this section.
 - ⁶ Panel D of Table A1 in Appendix A summarises the studies analysed in this section.
 - ⁷ Panel E of Table A1 in Appendix A summarises the papers analysed in this section.
 - ⁸ Panel F of Table A1 in Appendix A summarises the works reviewed in this section.
 - ⁹ Panel G of Table A1 in Appendix A summarises the papers scrutinised in this section.
 - ¹⁰ Panel H of Table A1 in Appendix A summarises the studies analysed in this section.
 - ¹¹ Panel I of Table A1 in Appendix A summarises the papers reviewed in this section.

References

- Abhyankar, Abhay, Bing Xu, and Jiayue Wang. 2013. Oil Price Shocks and the Stock Market: Evidence from Japan. *The Energy Journal* 34: 199–222. [CrossRef]
- Adekoya, Oluwasegun B., and Johnson A. Oliyide. 2021. How COVID-19 drives connectedness among commodity and financial markets: Evidence from TVP-VAR and causality-in-quantiles techniques. *Resources Policy* 70: 101898. [CrossRef] [PubMed]
- Adekoya, Oluwasegun B., Johnson A. Oliyide, OlaOluwa S. Yaya, and Mamdouh Abdulaziz Saleh Al-Faryan. 2022. Does oil connect differently with prominent assets during war? Analysis of intra-day data during the Russia-Ukraine saga. *Resources Policy* 77: 102728. [CrossRef]
- Aggarwal, Raj, Aigbe Akhigbe, and Sunil K. Mohanty. 2012. Oil price shocks and transportation firm asset prices. *Energy Economics* 34: 1370–79. [CrossRef]
- Ajmi, Ahdi Noomen, Ghassen El-montasser, Shawkat Hammoudeh, and Duc Khuong Nguyen. 2014. Oil prices and MENA stock markets: New evidence from nonlinear and asymmetric causalities during and after the crisis period. *Applied Economics* 46: 2167–77. [CrossRef]
- Akyildirim, Erdinc, Oguzhan Cepni, Peter Molnar, and Gazi Salah Uddin. 2022. Connectedness of energy markets around the world during the COVID-19 pandemic. *Energy Economics* 109: 105900. [CrossRef] [PubMed]
- Aloui, Chaker, and Rania Jammazi. 2009. The effects of crude oil shocks on stock market shifts behaviour: A regime switching approach. *Energy Economics* 31: 789–99. [CrossRef]
- Alsaman, Zeina. 2016. Oil price uncertainty and the U.S. stock market analysis based on a GARCH-in-mean VAR model. *Energy Economics* 59: 251–60. [CrossRef]
- Al-Yahyaee, Khamis Hamed, Walid Mensi, Ahmet Sensoy, and Sang Hoon Kang. 2019. Energy, precious metals, and GCC stock markets: Is there any risk spillover? *Pacific-Basin Finance Journal* 56: 45–70. [CrossRef]
- Angelidis, Timotheos, Stavros Degiannakis, and George Filis. 2015. US stock market regimes and oil price shocks. *Global Finance Journal* 28: 132–46. [CrossRef]
- Antonakakis, Nikolaos, and George Filis. 2013. Oil Prices and Stock Market Correlation: A Time-Varying Approach. *International Journal of Energy and Statistics* 1: 17–29. [CrossRef]
- Antonakakis, Nikolaos, Juncal Cunado, George Filis, David Gabauer, and Fernando Perez de Gracia. 2018. Oil volatility, oil and gas firms and portfolio diversification. *Energy Economics* 70: 499–515. [CrossRef]
- Apergis, Nicholas, and Stephen M. Miller. 2009. Do structural oil-market shocks affect stock prices? *Energy Economics* 31: 569–75. [CrossRef]
- Arouri, Mohamed El Hedi. 2011. Does crude oil move stock markets in Europe? A sector investigation. *Economic Modelling* 28: 1716–25. [CrossRef]
- Arouri, Mohamed El Hedi, Amine Lahiani, and Duc Khuong Nguyen. 2011a. Return and volatility transmission between world oil prices and stock markets of the GCC countries. *Economic Modelling* 28: 1815–25. [CrossRef]
- Arouri, Mohamed El Hedi, and Christophe Rault. 2012. Oil prices and stock markets in GCC countries: Empirical evidence from panel analysis. *International Journal of Finance and Economics* 17: 242–53. [CrossRef]
- Arouri, Mohamed El Hedi, Jamel Jouini, and Duc Khuong Nguyen. 2011b. Volatility spillovers between oil prices and stock sector returns: Implications for portfolio management. *Journal of International Money and Finance* 30: 1387–405. [CrossRef]
- Arouri, Mohamed El Hedi, Jamel Jouini, and Duc Khuong Nguyen. 2012. On the impacts of oil price fluctuations on European equity markets: Volatility spillover and hedging effectiveness. *Energy Economics* 34: 611–17. [CrossRef]
- Ashfaq, Saleha, Yong Tang, and Rashid Maqbool. 2019. Volatility spillover impact of world oil prices on leading Asian energy exporting and importing economies' stock returns. *Energy* 188: 116002. [CrossRef]
- Asteriou, Dimitrios, and Yuliya Bashmakova. 2013. Assessing the impact of oil returns on emerging stock markets: A panel data approach for ten Central and Eastern European Countries. *Energy Economics* 38: 204–11. [CrossRef]
- Awartani, Basel, and Aktham Issa Maghyreh. 2013. Dynamic spillovers between oil and stock markets in the Gulf Cooperation Council Countries. *Energy Economics* 36: 28–42. [CrossRef]
- Bagirov, Miramir, and Cesario Mateus. 2019. Oil prices, stock markets and firm performance: Evidence from Europe. *International Review of Economics and Finance* 61: 270–88. [CrossRef]
- Bagirov, Miramir, and Cesario Mateus. 2022. Petroleum prices and equity sector returns in petroleum exporting and importing countries: An analysis of volatility transmissions and hedging. *Applied Economics* 54: 2610–26. [CrossRef]
- Barndorff-Nielsen, Ole E., Silja Kinnebrok, and Neil Shephard. 2010. Measuring downside risk-realised semivariance. In *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*. Edited by Tim Bollerslev, Jeffrey Russell and Mark Watson. Oxford: Oxford University Press, pp. 117–36.
- Basher, Syed Abul, Alfred A. Haug, and Perry Sadorsky. 2012. Oil prices, exchange rates and emerging stock markets. *Energy Economics* 34: 227–40. [CrossRef]
- Basher, Syed Abul, Alfred A. Haug, and Perry Sadorsky. 2018. The impact of oil-market shocks on stock returns in major oil-exporting countries. *Journal of International Money and Finance* 86: 264–80. [CrossRef]
- Basher, Syed Abul, and Perry Sadorsky. 2006. Oil price risk and emerging stock markets. *Global Finance Journal* 17: 224–51. [CrossRef]
- Bastianin, Andrea, Francesca Conti, and Matteo Manera. 2016. The impacts of oil price shocks on stock market volatility: Evidence from the G7 countries. *Energy Policy* 98: 160–69. [CrossRef]

- Belhassine, Olfa. 2020. Volatility spillovers and hedging effectiveness between the oil market and Eurozone sectors: A tale of two crises. *Research in International Business and Finance* 53: 101195. [CrossRef]
- Belhassine, Olfa, and Chiraz Karamti. 2021. Volatility spillovers and hedging effectiveness between oil and stock markets: Evidence from a wavelet-based and structural breaks analysis. *Energy Economics* 102: 105513. [CrossRef]
- Benlagha, Nouredine, and Salaheddine El Omari. 2022. Connectedness of stock markets with gold and oil: New evidence from COVID-19 pandemic. *Finance Research Letters* 46: 102373. [CrossRef]
- Bhar, Ramaprasad, and Biljana Nikolova. 2010. Global oil prices, oil industry and equity returns: Russian experience. *Scottish Journal of Political Economy* 57: 169–86. [CrossRef]
- Bjornland, Hilde C. 2000. The dynamic effects of aggregate demand, supply and oil price shocks—A comparative study. *The Manchester School of Economic Studies* 68: 578–607. [CrossRef]
- Bjornland, Hilde C. 2009. Oil price shocks and stock market booms in an oil exporting country. *Scottish Journal of Political Economy* 56: 232–54. [CrossRef]
- Boldanov, Rustam, Stavros Degiannakis, and George Filis. 2016. Time-varying correlation between oil and stock market volatilities: Evidence from oil-importing and oil-exporting countries. *International Review of Financial Analysis* 48: 209–20. [CrossRef]
- Bouri, Elie. 2015. Oil volatility shocks and the stock markets of oil-importing MENA economies: A tale from the financial crisis. *Energy Economics* 51: 590–98. [CrossRef]
- Bouri, Elie, Basel Awartani, and Aktham Maghyereh. 2016. Crude oil prices and sectoral stock returns in Jordan around the Arab uprisings of 2010. *Energy Economics* 56: 205–14. [CrossRef]
- Bouri, Elie, Oguzhan Cepni, David Gabauer, and Rangan Gupta. 2021. Return connectedness across asset classes around the COVID-19 outbreak. *International Review of Financial Analysis* 73: 101646. [CrossRef]
- Boyer, M. Martin, and Didier Filion. 2007. Common and fundamental factors in stock returns of Canadian oil and gas companies. *Energy Economics* 29: 428–53. [CrossRef]
- British Petroleum. 2022. *BP Statistical Review of World Energy 2022*. London: BP.
- Broadstock, David C., and George Filis. 2014. Oil price shocks and stock market returns: New evidence from the United States and China. *Journal of International Financial Markets, Institutions & Money* 33: 417–33.
- Broadstock, David C., Hong Cao, and Dayong Zhang. 2012. Oil shocks and their impact on energy related stocks in China. *Energy Economics* 34: 1888–95. [CrossRef]
- Broadstock, David C., Rui Wang, and Dayong Zhang. 2014. Direct and indirect oil shocks and their impacts upon energy related stocks. *Energy Systems* 38: 451–67. [CrossRef]
- Burbidge, John, and Alan Harrison. 1984. Testing for the effects of oil-price rises using vector autoregressions. *International Economic Review* 25: 459–84. [CrossRef]
- Cameron, Ken, and Oliver Schnusenberg. 2009. Oil prices, SUVs, and Iraq: An investigation of automobile manufacturer oil price sensitivity. *Energy Economics* 31: 375–81. [CrossRef]
- Cashin, Paul, Kamiar Mohaddes, Maziar Raissi, and Mehdi Raissi. 2014. The differential effects of oil demand and supply shocks on the global economy. *Energy Economics* 44: 113–34. [CrossRef]
- Chang, Chia-Lin, Michael McAleer, and Roengchai Tansuchat. 2013. Conditional correlations and volatility spillovers between crude oil and stock index returns. *North American Journal of Economics and Finance* 25: 116–38. [CrossRef]
- Chang, Kuang-Liang, and Shih-Ti Yu. 2013. Does crude oil price play an important role in explaining stock return behavior? *Energy Economics* 39: 159–68. [CrossRef]
- Chen, Shiu-Sheng. 2010. Do higher oil prices push the stock market into bear territory? *Energy Economics* 32: 490–95. [CrossRef]
- Ciner, Cetin. 2013. Oil and stock returns: Frequency domain evidence. *Journal of International Financial Markets, Institutions & Money* 23: 1–11.
- Cong, Rong-Gang, Yi-Ming Wei, Jian-Lin Jiao, and Ying Fan. 2008. Relationships between oil price shocks and stock market: An empirical analysis from China. *Energy Policy* 36: 3544–53. [CrossRef]
- Costola, Michele, and Marco Lorusso. 2022. Spillovers among energy commodities and the Russian stock market. *Journal of Commodity Markets* 28: 100249. [CrossRef]
- Cunado, Juncal, and Fernando Perez de Gracia. 2003. Do oil price shocks matter? Evidence for some European countries. *Energy Economics* 25: 137–54. [CrossRef]
- Cunado, Juncal, and Fernando Perez de Gracia. 2005. Oil prices, economic activity and inflation: Evidence for some Asian countries. *The Quarterly Review of Economics and Finance* 45: 65–83. [CrossRef]
- Cunado, Juncal, and Fernando Perez de Gracia. 2014. Oil price shocks and stock market returns: Evidence for some European countries. *Energy Economics* 42: 365–77. [CrossRef]
- Dayanandan, Ajit, and Han Donker. 2011. Oil prices and accounting profits of oil and gas companies. *International Review of Financial Analysis* 20: 252–57. [CrossRef]
- Degiannakis, Stavros, George Filis, and Christos Floros. 2013. Oil and stock returns: Evidence from European industrial sector indices in a time-varying environment. *Journal of International Financial Markets, Institutions & Money* 26: 175–91.
- Degiannakis, Stavros, George Filis, and Renatas Kizys. 2014. The Effects of Oil Price Shocks on Stock Market Volatility: Evidence from European Data. *The Energy Journal* 35: 35–56. [CrossRef]

- Degiannakis, Stavros, George Filis, and Vipin Arora. 2018. Oil Prices and Stock Markets: A Review of the Theory and Empirical Evidence. *The Energy Journal* 39: 85–130. [CrossRef]
- Demirer, Riza, Roman Ferrer, and Syed Jawad Hussain Shahzad. 2020. Oil price shocks, global financial markets and their connectedness. *Energy Economics* 88: 104771. [CrossRef]
- Demirer, Riza, Shrikant P. Jategaonkar, and Ahmed A. A. Khalifa. 2015. Oil price risk exposure and the cross-section of stock returns: The case of net exporting countries. *Energy Economics* 49: 132–40. [CrossRef]
- Driesprong, Gerben, Ben Jacobsen, and Benjamin Maat. 2008. Striking oil: Another puzzle? *Journal of Financial Economics* 89: 307–27. [CrossRef]
- El-Sharif, Idris, Dick Brown, Bruce Burton, Bill Nixon, and Alex Russell. 2005. Evidence on the nature and extent of the relationship between oil prices and equity values in the UK. *Energy Economics* 27: 819–30. [CrossRef]
- Elyasiani, Elyas, Iqbal Mansur, and Babatunde Odusami. 2011. Oil price shocks and industry stock returns. *Energy Economics* 33: 966–74. [CrossRef]
- Enwereuzoh, Precious Adaku, Jones Odei-Mensah, and Peterson Owusu Junior. 2021. Crude oil shocks and African stock markets. *Research in International Business and Finance* 55: 101346. [CrossRef]
- Ewing, Bradley T., and Farooq Malik. 2016. Volatility spillovers between oil prices and the stock market under structural breaks. *Global Finance Journal* 29: 12–23. [CrossRef]
- Faff, Robert W., and Timothy J. Brailsford. 1999. Oil price risk and the Australian stock market. *Journal of Energy Finance and Development* 4: 69–87. [CrossRef]
- Fang, Chung-Rou, and Shih-Yi You. 2014. The impact of oil price shocks on the large emerging countries' stock prices: Evidence from China, India and Russia. *International Review of Economics and Finance* 29: 330–38. [CrossRef]
- Farid, Saqib, Ghulam Mujtaba Kayani, Muhammad Abubakr Naem, and Syed Jawad Hussain Shahzad. 2021. Intraday volatility transmission among precious metals, energy and stocks during the COVID-19 pandemic. *Resources Policy* 72: 102101. [CrossRef]
- Filis, George. 2010. Macro economy, stock market and oil prices: Do meaningful relationships exist among their cyclical fluctuations? *Energy Economics* 32: 877–86. [CrossRef]
- Filis, George, and Ioannis Chatziantoniou. 2014. Financial and monetary policy responses to oil price shocks: Evidence from oil-importing and oil-exporting countries. *Review of Quantitative Finance and Accounting* 42: 709–29. [CrossRef]
- Filis, George, Stavros Degiannakis, and Christos Floros. 2011. Dynamic correlation between stock market and oil prices: The case of oil-importing and oil-exporting countries. *International Review of Financial Analysis* 20: 152–64. [CrossRef]
- Gil-Alana, Luis A., and Olaoluwa S. Yaya. 2014. The relationship between oil prices and the Nigerian stock market. An analysis based on fractional integration and cointegration. *Energy Economics* 46: 328–33. [CrossRef]
- Gisser, Micha, and Thomas H. Goodwin. 1986. Crude oil and the macroeconomy: Tests of some popular notions. *Journal of Money, Credit and Banking* 18: 95–103. [CrossRef]
- Gogineni, Sridhar. 2010. Oil and the Stock Market: An Industry Level Analysis. *The Financial Review* 45: 995–1010. [CrossRef]
- Guesmi, Khaled, and Salma Fattoum. 2014. Return and volatility transmission between oil prices and oil-exporting and oil-importing countries. *Economic Modelling* 38: 305–10. [CrossRef]
- Guntner, Jochen H. F. 2014. How do international stock markets respond to oil demand and supply shocks? *Macroeconomic Dynamics* 18: 1657–82. [CrossRef]
- Gupta, Eshita. 2008. Oil vulnerability index of oil-importing countries. *Energy Policy* 36: 1195–211. [CrossRef]
- Gupta, Kartick. 2016. Oil price shocks, competition, and oil & gas stock returns—Global evidence. *Energy Economics* 57: 140–53.
- Gupta, Rangan, and Mampho P. Modise. 2013. Does the source of oil price shocks matter for South African stock returns? A structural VAR approach. *Energy Economics* 40: 825–31. [CrossRef]
- Hamilton, James D. 1983. Oil and the macro-economy since World War II. *Journal of Political Economy* 91: 228–48. [CrossRef]
- Hamilton, James D. 1996. This is what happened to the oil price-macroeconomy relationship. *Journal of Monetary Economics* 38: 215–20. [CrossRef]
- Hamilton, James D. 2003. What is an oil shock? *Journal of Econometrics* 113: 363–98. [CrossRef]
- Hammoudeh, Shawkat, and Huimin Li. 2005. Oil sensitivity and systematic risk in oil-sensitive stock indices. *Journal of Economics and Business* 57: 1–21. [CrossRef]
- Hammoudeh, Shawkat, Sel Dibooglu, and Eisa Aleisa. 2004. Relationships among U.S. oil prices and oil industry equity indices. *International Review of Economics and Finance* 13: 427–53. [CrossRef]
- Hatemi, J. Abdunasser, Abdulrahman Al Shayeb, and Eduardo Roca. 2017. The effect of oil prices on stock prices: Fresh evidence from asymmetric causality tests. *Applied Economics* 49: 1584–92. [CrossRef]
- Heinlein, Reinhold, Gabriella D. Legrenzi, and Scott M. R. Mahadeo. 2021. Crude oil and stock markets in the COVID-19 crisis: Evidence from oil exporters and importers. *The Quarterly Review of Economics and Finance* 82: 223–29. [CrossRef]
- Henriques, Irene, and Perry Sadorsky. 2008. Oil prices and the stock prices of alternative energy companies. *Energy Economics* 30: 998–1010. [CrossRef]
- Hooker, Mark A. 1996. What happened to the oil price-macroeconomy relationship? *Journal of Monetary Economics* 38: 195–213. [CrossRef]
- Hou, Keqiang, Dean C. Mountain, and Ting Wu. 2016. Oil price shocks and their transmission mechanism in an oil-exporting economy: A VAR analysis informed by a DSGE model. *Journal of International Money and Finance* 68: 21–49. [CrossRef]

- Huang, Roger D., Ronald W. Masulis, and Hans R. Stoll. 1996. Energy shocks and financial markets. *The Journal of Futures Markets* 16: 1–27. [CrossRef]
- Hung, Ngo Thai, and Xuan Vinh Vo. 2021. Directional spillover effects and time-frequency nexus between oil, gold and stock markets: Evidence from pre and during COVID-19 outbreak. *International Review of Financial Analysis* 76: 101730. [CrossRef]
- Hwang, Inwook, and Jaebom Kim. 2021. Oil price shocks and the US stock market: A nonlinear approach. *Journal of Empirical Finance* 64: 23–36. [CrossRef]
- Jammazi, Rania, and Chaker Aloui. 2010. Wavelet decomposition and regime shifts: Assessing the effects of crude oil shocks on stock market returns. *Energy Policy* 38: 1415–35. [CrossRef]
- Jebabli, Ikram, Noureddine Kouaissah, and Mohamed Arouri. 2022. Volatility Spillovers between Stock and Energy Markets during Crises: A Comparative Assessment between the 2008 Global Financial Crisis and the COVID-19 Pandemic Crisis. *Finance Research Letters* 46: 102363. [CrossRef]
- Jimenez-Rodriguez, Rebeca, and Marcelo Sanchez. 2005. Oil price shocks and real GDP growth: Empirical evidence for some OECD countries. *Applied Economics* 37: 201–28. [CrossRef]
- Jones, Charles M., and Gautam Kaul. 1996. Oil and the stock markets. *Journal of Finance* 51: 463–91. [CrossRef]
- Kang, Wensheng, and Ronald A. Ratti. 2013. Oil shocks, policy uncertainty and stock market return. *Journal of International Financial Markets, Institutions & Money* 26: 305–18.
- Kang, Wensheng, Fernando Perez de Gracia, and Ronald A. Ratti. 2017. Oil price shocks, policy uncertainty, and stock returns of oil and gas corporations. *Journal of International Money and Finance* 70: 344–59. [CrossRef]
- Kang, Wensheng, Ronald A. Ratti, and Joaquin Vespignani. 2016. The impact of oil price shocks on the U.S. stock market: A note on the roles of U.S. and non-U.S. oil production. *Economic Letters* 145: 176–81. [CrossRef]
- Kang, Wensheng, Ronald A. Ratti, and Kyung Hwan Yoon. 2015. The impact of oil price shocks on the stock market return and volatility relationship. *Journal of International Financial Markets, Institutions & Money* 34: 41–54.
- Khalfaoui, Rabeh, Mohamed Boutahar, and Heni Boubaker. 2015. Analyzing volatility spillovers and hedging between oil and stock markets: Evidence from wavelet analysis. *Energy Economics* 49: 540–49. [CrossRef]
- Kilian, Lutz. 2009. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *The American Economic Review* 99: 1053–69. [CrossRef]
- Kilian, Lutz, and Cheolbeom Park. 2009. The impact of oil price shocks on the US stock market. *International Economic Review* 50: 1267–87. [CrossRef]
- Kilian, Lutz, and Daniel P. Murphy. 2014. The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied Econometrics* 29: 454–78. [CrossRef]
- Kim, Won Joong, Shawkat Hammoudeh, Jun Seog Hyun, and Rangan Gupta. 2017. Oil price shocks and China's economy: Reactions of the monetary policy to oil price shocks. *Energy Economics* 62: 61–69. [CrossRef]
- Koh, Wee Chian. 2017. How do oil supply and demand shocks affect Asian stock markets? *Macroeconomics and Finance in Emerging Market Economies* 10: 1–18. [CrossRef]
- Korhonen, Iikka, and Svetlana Ledyaeva. 2010. Trade linkages and macroeconomic effects of the price of oil. *Energy Economics* 32: 848–56. [CrossRef]
- Kristjanpoller, Werner D., and Diego Concha. 2016. Impact of fuel price fluctuations on airline stock returns. *Applied Energy* 178: 496–504. [CrossRef]
- Kumar, Surender, Shunsuke Managi, and Akimi Matsuda. 2012. Stock prices of clean energy firms, oil and carbon markets: A vector autoregressive analysis. *Energy Economics* 34: 215–26. [CrossRef]
- Lardic, Sandrine, and Valerie Mignon. 2006. The impact of oil prices on GDP in European countries: An empirical investigation based on asymmetric cointegration. *Energy Policy* 34: 3910–15. [CrossRef]
- Lardic, Sandrine, and Valerie Mignon. 2008. Oil prices and economic activity: An asymmetric cointegration approach. *Energy Economics* 30: 847–55. [CrossRef]
- Le, Thai-Ha, and Youngho Chang. 2015. Effects of oil price shocks on the stock market performance: Do nature of shocks and economies matter? *Energy Economics* 51: 261–74. [CrossRef]
- Lee, Bi-Juan, Chin Wei Yang, and Bwo-Nung Huang. 2012. Oil price movements and stock markets revisited: A case of sector stock price indexes in the G-7 countries. *Energy Economics* 34: 1284–300. [CrossRef]
- Lee, Chi-Chuan, Chien-Chiang Lee, and Shao-Lin Ning. 2017. Dynamic relationship of oil price shocks and country risks. *Energy Economics* 66: 571–81. [CrossRef]
- Li, Qiming, Ke Cheng, and Xiaoguang Yang. 2017. Response pattern of stock returns to international oil price shocks: From the perspective of China's oil industrial chain. *Applied Energy* 185: 1821–31. [CrossRef]
- Li, Su-Fang, Hui-Ming Zhu, and Keming Yu. 2012. Oil prices and stock market in China: A sector analysis using panel cointegration with multiple breaks. *Energy Economics* 34: 1951–58. [CrossRef]
- Lin, Boqiang, Presley K. Wesseh, and Michael Owusu Appiah. 2014. Oil price fluctuation, volatility spillover and the Ghanaian equity market: Implication for portfolio management and hedging effectiveness. *Energy Economics* 42: 172–82. [CrossRef]
- Lin, Jie, Hao Xiao, and Jian Chai. 2023. Dynamic effects and driving intermediations of oil price shocks on major economies. *Energy Economics* 124: 106779. [CrossRef]

- Liu, Donghui, Lingjie Meng, and Yudong Wang. 2020. Oil price shocks and Chinese economy revisited: New evidence from SVAR model with sign restrictions. *International Review of Economics & Finance* 69: 20–32.
- Lorusso, Marco, and Luca Pieroni. 2018. Causes and consequences of oil price shocks on the UK economy. *Economic Modelling* 72: 223–36. [CrossRef]
- Maghyereh, Aktham I., Basel Awartani, and Elie Bouri. 2016. The directional volatility connectedness between crude oil and equity markets: New evidence from implied volatility indexes. *Energy Economics* 57: 78–93. [CrossRef]
- Malik, Farooq, and Bradley T. Ewing. 2009. Volatility transmission between oil prices and equity sector returns. *International Review of Financial Analysis* 18: 95–100. [CrossRef]
- Malik, Farooq, and Shawkat Hammoudeh. 2007. Shock and volatility transmission in the oil, US and Gulf equity markets. *International Review of Economics and Finance* 16: 357–68. [CrossRef]
- Martin-Barragan, Belen, Sofia B. Ramos, and Helena Veiga. 2015. Correlations between oil and stock markets: A wavelet-based approach. *Economic Modelling* 50: 212–27. [CrossRef]
- Masih, Rumi, Sanjay Peters, and Lurion De Mello. 2011. Oil price volatility and stock price fluctuations in an emerging market: Evidence from South Korea. *Energy Economics* 33: 975–86. [CrossRef]
- Mateus, Cesario, Raju Chinthalapati, and Irina B. Mateus. 2017. Intraday industry-specific spillover effect in European equity markets. *The Quarterly Review of Economics and Finance* 63: 278–98. [CrossRef]
- Mensi, Walid, Makram Beljid, Adel Boubaker, and Shunsuke Managi. 2013. Correlations and volatility spillovers across commodity and stock markets: Linking energies, food, and gold. *Economic Modelling* 32: 15–22. [CrossRef]
- Mensi, Walid, Muhammad Shafiullah, Xuan Vinh Vo, and Sang Hoon Kang. 2021. Volatility spillovers between strategic commodity futures and stock markets and portfolio implications: Evidence from developed and emerging economies. *Resources Policy* 71: 102002. [CrossRef]
- Mensi, Walid, Xuan Vinh Vo, and Sang Hoon Kang. 2022. COVID-19 pandemic's impact on intraday volatility spillover between oil, gold, and stock markets. *Economic Analysis and Policy* 74: 702–15. [CrossRef]
- Miller, J. Isaac, and Ronald A. Ratti. 2009. Crude oil and stock markets: Stability, instability and bubbles. *Energy Economics* 31: 559–68. [CrossRef]
- Mishra, Shekhar, and Sibanjan Mishra. 2021. Are Indian sectoral indices oil shock prone? An empirical evaluation. *Resources Policy* 70: 101889. [CrossRef]
- Mohanty, Sunil K., Aigbe Akhigbe, Tawfeek A. Al-Khyal, and Turki Bugshan. 2013. Oil and stock market activity when prices go up and down: The case of the oil and gas industry. *Review of Quantitative Finance and Accounting* 41: 253–72. [CrossRef]
- Mohanty, Sunil K., and Mohan Nandha. 2011. Oil risk exposure: The case of the U.S. oil and gas sector. *The Financial Review* 46: 165–91. [CrossRef]
- Mohanty, Sunil K., Mohan Nandha, Abdullah Q. Turkistani, and Muhammed Y. Alaitani. 2011. Oil price movements and stock market returns: Evidence from Gulf Cooperation Council (GCC) countries. *Global Finance Journal* 22: 42–55. [CrossRef]
- Mohanty, Sunil, Mohan Nandha, and Gabor Bota. 2010. Oil shocks and stock returns: The case of the Central and Eastern European (CEE) oil and gas sectors. *Emerging Markets Review* 11: 358–72. [CrossRef]
- Mokni, Khalel. 2020. Time-varying effect of oil price shocks on the stock market returns: Evidence from oil-importing and oil-exporting countries. *Energy Reports* 6: 605–19. [CrossRef]
- Mollick, Andre Varella, and Tibebe Abebe Assefa. 2013. U.S. stock returns and oil prices: The tale from daily data and the 2008–2009 financial crisis. *Energy Economics* 36: 1–18. [CrossRef]
- Mork, Knut Anton. 1989. Oil and the macroeconomy when prices go up and down: An extension of Hamilton's results. *Journal of Political Economy* 97: 740–44. [CrossRef]
- Mork, Knut Anton, Oystein Olsen, and Hans Terje Mysen. 1994. Macroeconomic responses to oil price increases and decreases in seven OECD countries. *Energy Journal* 15: 19–35. [CrossRef]
- Moya-Martinez, Pablo, Roman Ferrer-Lapena, and Francisco Escribano-Sotos. 2014. Oil price risk in the Spanish stock market: An industry perspective. *Economic Modelling* 37: 280–90. [CrossRef]
- Nandha, Mohan, and Robert Brooks. 2009. Oil prices and transport sector returns: An international analysis. *Review of Quantitative Finance and Accounting* 33: 393–409. [CrossRef]
- Nandha, Mohan, and Robert Faff. 2008. Does oil move equity prices? A global view. *Energy Economics* 30: 986–97. [CrossRef]
- Narayan, Paresh Kumar, and Rangan Gupta. 2015. Has oil price predicted stock returns for over a century? *Energy Economics* 48: 18–23. [CrossRef]
- Narayan, Paresh Kumar, and Seema Narayan. 2010. Modelling the impact of oil prices on Vietnam's stock prices. *Applied Energy* 87: 356–61. [CrossRef]
- Narayan, Paresh Kumar, and Susan Sunila Sharma. 2011. New evidence on oil price and firm returns. *Journal of Banking & Finance* 35: 3253–3262.
- Narayan, Paresh Kumar, and Susan Sunila Sharma. 2014. Firm return volatility and economic gains: The role of oil prices. *Economic Modelling* 38: 142–51. [CrossRef]
- Nusair, Salah A. 2016. The effects of oil price shocks on the economies of the Gulf Co-operation Council countries: Nonlinear analysis. *Energy Policy* 91: 256–67. [CrossRef]

- O'Neill, Terence J., Jack Penm, and Richard D. Terrell. 2008. The role of higher oil prices: A case of major developed countries. *Research in Finance* 24: 287–99.
- Papapetrou, Evangelia. 2001. Oil price shocks, stock market, economic activity an employment in Greece. *Energy Economics* 23: 511–32. [CrossRef]
- Park, J., and Ronald A. Ratti. 2008. Oil price shocks and stock markets in the U.S. and 13 European countries. *Energy Economics* 30: 2587–2608. [CrossRef]
- Peersman, Gert, and Ine Van Robays. 2012. Cross-country differences in the effects of oil shocks. *Energy Economics* 34: 1532–1547. [CrossRef]
- Phan, Dinh Hoang Bach, Susan Sunila Sharma, and Paresh Kumar Narayan. 2015. Oil price and stock returns of consumers and producers of crude oil. *Journal of International Financial Markets, Institutions & Money* 34: 245–62.
- Ramos, Sofia B., and Helena Veiga. 2011. Risk factors in oil and gas industry returns: International evidence. *Energy Economics* 33: 525–42. [CrossRef]
- Ramos, Sofia B., and Helena Veiga. 2013. Oil price asymmetric effects: Answering the puzzle in international stock markets. *Energy Economics* 38: 136–45. [CrossRef]
- Ready, Robert C. 2018. Oil Prices and the Stock Market. *Review of Finance* 22: 155–76. [CrossRef]
- Reboredo, Juan C. 2010. Nonlinear effects of oil shocks on stock returns: A Markov-switching approach. *Applied Economics* 42: 3735–44. [CrossRef]
- Reboredo, Juan C., and Andrea Ugolini. 2016. Quantile dependence of oil price movements and stock returns. *Energy Economics* 54: 33–49. [CrossRef]
- Reboredo, Juan C., and Miguel A. Rivera-Castro. 2014. Wavelet-based evidence of the impact of oil prices on stock returns. *International Review of Economics and Finance* 29: 145–76. [CrossRef]
- Sadorsky, Perry. 1999. Oil price shocks and stock market activity. *Energy Economics* 21: 449–69. [CrossRef]
- Sadorsky, Perry. 2001. Risk factors in stock returns of Canadian oil and gas companies. *Energy Economics* 23: 17–28. [CrossRef]
- Sadorsky, Perry. 2008. Assessing the impact of oil prices on firms of different sizes: Its tough being in the middle. *Energy Policy* 36: 3854–61. [CrossRef]
- Sadorsky, Perry. 2012. Correlations and volatility spillovers between oil prices and the stock prices of clean energy and technology companies. *Energy Economics* 34: 248–55. [CrossRef]
- Sakaki, Hamid. 2019. Oil price shocks and the equity market: Evidence for the S&P 500 sectoral indices. *Research in International Business and Finance* 49: 137–55.
- Salisu, Afees A., and Kazeem O. Isah. 2017. Revisiting the oil price and stock market nexus: A nonlinear Panel ARDL approach. *Economic Modelling* 66: 258–71. [CrossRef]
- Salisu, Afees A., and Tirimisiyu F. Oloko. 2015. Modeling oil price–US stock nexus: A VARMA–BEKK–AGARCH approach. *Energy Economics* 50: 1–12. [CrossRef]
- Sarwar, Suleman, Aviral Kumar Tiwari, and Cao Tingqiu. 2020. Analyzing volatility spillovers between oil market and Asian stock markets. *Resources Policy* 66: 101608. [CrossRef]
- Sarwar, Suleman, Rabeh Khalfaoui, Rida Waheed, and Hamidreza Ghorbani Dastgerdi. 2019. Volatility spillovers and hedging: Evidence from Asian oil-importing countries. *Resources Policy* 61: 479–88. [CrossRef]
- Scholtens, Bert, and Cenk Yurtsever. 2012. Oil price shocks and European industries. *Energy Economics* 34: 1187–95. [CrossRef]
- Silvapulle, Param, Russell Smyth, Xibin Zhang, and Jean-Pierre Fenech. 2017. Nonparametric panel data model for crude oil and stock market prices in net oil importing countries. *Energy Economics* 67: 255–67. [CrossRef]
- Silvennoinen, Annastiina, and Susan Thorp. 2013. Financialization, crisis and commodity correlation dynamics. *Journal of International Financial Markets, Institutions & Money* 24: 42–65.
- Smiech, Slawomir, Monika Papiez, Michal Rubaszek, and Malgorzata Snarska. 2021. The role of oil price uncertainty shocks on oil-exporting countries. *Energy Economics* 93: 105028. [CrossRef]
- Smyth, Russell, and Paresh Kumar Narayan. 2018. What do we know about oil prices and stock returns? *International Review of Financial Analysis* 57: 148–56. [CrossRef]
- Suleman, Muhammad Tahir, Ron McIver, and Sang Hoon Kang. 2021. Asymmetric volatility connectedness between Islamic stock and commodity markets. *Global Finance Journal* 49: 100653. [CrossRef]
- Tan, Yan, and Utai Uprasen. 2023. Asymmetric effects of oil price shocks on income inequality in ASEAN countries. *Energy Economics* 126: 107033. [CrossRef]
- Tang, Ke, and Wei Xiong. 2012. Index investment and the financialization of commodities. *Financial Analysts Journal* 68: 54–74. [CrossRef]
- Tiwari, Aviral Kumar, Samia Nasreen, Subhan Ullah, and Muhammad Shahbaz. 2021. Analysing spillover between returns and volatility series of oil across major stock markets. *International Journal of Finance and Economics* 26: 2458–90. [CrossRef]
- Tsai, Chun-Li. 2015. How do U.S. stock returns respond differently to oil price shocks pre-crisis, within the financial crisis, and post-crisis? *Energy Economics* 50: 47–62. [CrossRef]
- Umar, Zaghum, Francisco Jareno, and Ana Escribano. 2020. Static and dynamic connectedness between oil price shocks and Spanish equities: A sector analysis. *The European Journal of Finance* 27: 880–96. [CrossRef]

- Umar, Zaghum, Nader Trabelsi, and Adam Zaremba. 2021. Oil shocks and equity markets: The case of GCC and BRICS economies. *Energy Economics* 96: 105155. [CrossRef]
- Vo, Minh. 2011. Oil and stock market volatility: A multivariate stochastic volatility perspective. *Energy Economics* 33: 956–65. [CrossRef]
- Wang, Yudong, and Li Liu. 2016. Crude oil and world stock markets: Volatility spillovers, dynamic correlations, and hedging. *Empirical Economics* 50: 1481–509. [CrossRef]
- Wang, Yudong, Chongfeng Wu, and Li Yang. 2013. Oil price shocks and stock market activities: Evidence from oil-importing and oil-exporting countries. *Journal of Comparative Economics* 41: 1220–39. [CrossRef]
- Wen, Fenghua, Keli Zhang, and Xu Gong. 2021. The effects of oil price shocks on inflation in the G7 countries. *The North American Journal of Economics and Finance* 57: 101391. [CrossRef]
- Wong, Jin Boon. 2020. Stock market reactions to different types of oil shocks: Evidence from China. *Journal of Futures Markets* 41: 179–93. [CrossRef]
- Xu, Bing. 2015. Oil prices and UK industry-level stock returns. *Applied Economics* 47: 2608–27. [CrossRef]
- Xu, Weiju, Feng Ma, Wang Chen, and Bing Zhang. 2019. Asymmetric volatility spillovers between oil and stock markets: Evidence from China and the United States. *Energy Economics* 80: 310–20. [CrossRef]
- Yousaf, Imran, and Arshad Hassan. 2019. Linkages between crude oil and emerging Asian stock markets: New evidence from the Chinese stock market crash. *Finance Research Letters* 31: 207–17. [CrossRef]
- Zhang, Bing. 2017. How do great shocks influence the correlation between oil and international stock markets? *Applied Economics* 49: 1513–26. [CrossRef]
- Zhang, Hua, Jinyu Chen, and Liuguo Shao. 2021. Dynamic spillovers between energy and stock markets and their implications in the context of COVID-19. *International Review of Financial Analysis* 77: 101828. [CrossRef]
- Zhang, Wenting, and Shigeyuki Hamori. 2021. Crude oil market and stock markets during the COVID-19 pandemic: Evidence from the US, Japan, and Germany. *International Review of Financial Analysis* 74: 101702. [CrossRef]
- Zhu, Hui-Ming, Rong Li, and Sufang Li. 2014. Modelling dynamic dependence between crude oil prices and Asia-Pacific stock market returns. *International Review of Economics and Finance* 29: 208–23. [CrossRef]
- Zhu, Hui-Ming, Su-Fang Li, and Keming Yu. 2011. Crude oil shocks and stock markets: A panel threshold cointegration approach. *Energy Economics* 33: 987–94. [CrossRef]
- Zhu, Huiming, Xianfang Su, Wanhai You, and Yinghua Ren. 2017. Asymmetric effects of oil price shocks on stock returns: Evidence from a two-stage Markov regime-switching approach. *Applied Economics* 49: 2491–507. [CrossRef]
- Zhu, Huiming, Yawei Guo, Wanhai You, and Yaqin Xu. 2016. The heterogeneity dependence between crude oil price changes and industry stock market returns in China: Evidence from a quantile regression approach. *Energy Economics* 55: 30–41. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

The Financial Market of Indices of Socioeconomic Well-Being

Thilini V. Mahanama ¹, Abootaleb Shirvani ^{2,*}, Svetlozar Rachev ³ and Frank J. Fabozzi ⁴

¹ Department of Industrial Management, University of Kelaniya, Kelaniya 11600, Sri Lanka; thilininim@kln.ac.lk

² Department of Mathematical Science, Kean University, Union, NJ 07083, USA

³ Department of Mathematics & Statistics, Texas Tech University, Lubbock, TX 79409, USA; zari.rachev@ttu.edu

⁴ Carey Business School, Johns Hopkins University, Baltimore, MD 21218, USA; ffabozzi1@jhu.edu

* Correspondence: ashirvan@kean.edu

Abstract: This study discusses how financial economic theory and its quantitative tools can be applied to create socioeconomic indices and develop a financial market for the so-called “socioeconomic well-being indices”. In this study, we quantify socioeconomic well-being by assigning a dollar value to the well-being factors of selected countries; this is analogous to how the Dow 30 encapsulates the financial health of the US market. While environmental, social, and governance (ESG) financial markets address socioeconomic issues, our focus is broader, encompassing national citizens’ well-being. The dollar-denominated socioeconomic indices for each country can be viewed as financial assets that can serve as risky assets for constructing a global index, which, in turn, serves as a “market of well-being socioeconomic index”. This novel global index of well-being, paralleling the Dow Jones Industrial Average (DJIA), provides a comprehensive representation of the world’s socioeconomic status. Through advanced financial econometrics and dynamic asset pricing methodologies, we evaluate the potential for significant downturns in both the socioeconomic well-being indices of individual countries and the aggregate global index. This innovative approach allows us to engineer financial instruments akin to portfolio insurance, such as index puts, designed to hedge against these downturn risks. Our findings propose a financial market model for well-being indices, encouraging the financial industry to adopt and trade these indices as mechanisms to manage and hedge against downturn risks in well-being.

Keywords: socioeconomic well-being indices; dynamic asset pricing theory; world development socioeconomic indicators; global dollar socioeconomic well-being index; gross domestic product; financial econometric modeling

Citation: Mahanama, Thilini V., Abootaleb Shirvani, Svetlozar Rachev, and Frank J. Fabozzi. 2024. The Financial Market of Indices of Socioeconomic Well-Being. *Journal of Risk and Financial Management* 17: 35. <https://doi.org/10.3390/jrfm17010035>

Academic Editor: Thanasis Stengos

Received: 8 December 2023

Revised: 9 January 2024

Accepted: 10 January 2024

Published: 16 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A country’s policymakers require a measure of the magnitude of their economic output to evaluate the impact of economic policies on growth. Since 1934, the gross national product (GNP), proposed by Simon Kuznets, has served as the primary metric in the US and other developed countries, encompassing all final transactions of goods and services. In 1991, the gross domestic product (GDP), a similar measure, took the spotlight. However, the GNP/GDP faced criticism for focusing solely on aggregate economic wealth, neglecting individual citizens’ economic welfare and well-being. Leaders of major countries, like then-President of France Nicolas Sarkozy in 2008, recognized these limitations, leading to the identification of diverse sources of well-being beyond economic indicators that could be used to assess not just economic performance but also social progress.¹

The IMF’s report discusses income, consumption, and wealth distribution indicators for categorizing well-being and economic welfare (Quiros-Romero and Reinsdorf 2020). Ahmad and Qureshi (2021) highlight the need for an index measuring citizen well-being, allowing policymakers to benchmark their performance. This approach is exemplified by the United Nations Development Programme’s Human Development Index (HDI), which evaluates various dimensions of human development beyond economic growth (Sagar and

Najam 1998). The HDI represents achievements in health, knowledge, and living standards. The OECD has also pioneered economic indicators to comprehensively assess economic and social progress, such as the “Better Life Index”, which reflects a country’s current socioeconomic status based on factors like housing, income, education, and governance (OECD 2023).

Trindade et al. (2020) developed a socioeconomic well-being index using the historical mood of the US population, issuing marketable financial contracts like options and futures through financial econometric modeling and dynamic asset pricing theory. These indices offer early warnings for policymakers and private agents regarding potential mood-related market downturns. However, a country’s well-being is not solely dependent on the public mood or political factors; it should also encompass the role of the business community.

Our study develops a conceptual framework for encapsulating socioeconomic well-being; it is comparable to the way the Dow 30 indices encapsulate the industrial and financial vitality of corporations in the US. In this paper, we take a global view of a nation’s well-being, applying dynamic asset pricing theory and integrating financial economic theory to assess and manage the risks associated with socioeconomic indices, expanding on prior work on environment, society, and governance (ESG) and socially responsible investing (Matos 2020).

Our strategy involves establishing a financial market for socioeconomic well-being indices, treating them as risky financial assets, and assigning dollar values by normalizing these assets between 0 and 100; this is akin to ESG scores. We construct socioeconomic well-being indices for nine countries: the US, Australia, Brazil, China, Germany, India, Japan, South Africa, and the UK. Incorporating eight key world development indicators—the Gini index, unemployment, life expectancy at birth, gross national income, inflation rate, population, and foreign direct investment from the World Bank (IBRD 2022)—we create dollar-denominated indices, utilizing a country’s GDP to determine their financial value. These dollar-denominated indices are derived to measure and assign a tangible monetary value to the well-being factor, establishing a precedent in the financial analysis of social metrics.

We establish a market for dollar-denominated socioeconomic well-being indices, a *global index* akin to the Dow Jones Industrial Average (DJIA), treating them as risky financial assets. Our global index provides an equally weighted market index, with each country’s dollar-denominated socioeconomic well-being index considered a risky asset within this composite index. Leveraging these indices from developed countries, we conduct optimization analyses to derive efficient frontiers. We analyze the stability and risks associated with the socioeconomic well-being indices of individual nations and the global composite index by employing financial econometrics and dynamic asset pricing theory. We delve into the likelihood and potential impact of sharp downturns in these indices, which reflect broader socioeconomic challenges and trends.

By integrating financial theory with well-being metrics, we create a platform for financial instruments analogous to those used for hedging against market volatility, such as portfolio insurance. These instruments, such as index puts, are designed to mitigate the risk of downturns in the indices of well-being for countries and the global index alike. This concept opens up a new market for securities that protect against declines in social welfare, much like how traditional financial instruments protect against economic downturns. For financial markets, this represents an expansion of tradable assets to include indices that represent human and social capital. For policymakers and social scientists, it offers a new lens through which to view the economic implications of social policies and changes. For the global community, it emphasizes the importance of socioeconomic factors as integral elements of a nation’s wealth and prosperity.

Our analysis applies portfolio theory, complying with Basel II Accord requirements, and employs various risk–return measures. This paper discusses the development of a market where these socioeconomic well-being indices are traded as derivatives, providing a hedging mechanism against the socioeconomic risks faced by countries. The financial instruments that we propose are similar in nature to index puts, which allow investors to

insure their portfolios against declines. We anticipate that institutional investors will trade these assets similarly to other derivatives. For instance, a US equity portfolio manager forecasting a decline in the US dollar-denominated socioeconomic well-being index may opt for portfolio insurance instruments or short positions in the index. In the context of our socioeconomic well-being indices, these instruments would serve to protect against downturns in a nation's socioeconomic conditions. The availability of such instruments could also facilitate a more proactive approach to social investment and policymaking, encouraging a forward-looking stance on well-being.

Existing well-being indices lack dynamic features; they lack a time series econometric model that can be used to forecast and trade them in financial portfolios. No research has explored financial instruments for assessing a nation's downturns. Thus, we establish a market for socioeconomic indices, enabling the financial industry to engage in societal well-being assessment and the management of potential downturns. While ESG plays a vital role in addressing socioeconomic concerns in financial markets, our aim is broader, encompassing a nation's overall well-being.

During the preparation of this work, the authors utilized ChatGPT 3.5 and ChatGPT 4 for the sole purpose of enhancing the linguistic quality of our manuscript. After employing these AI tools, the authors meticulously reviewed and edited the content as needed, thereby taking full responsibility for the content of the publication. The use of these AI resources was strictly limited to improving the clarity and coherence of the text, given the diverse writing styles of the four authors. At no point was AI used for any analytical input or data analysis within the paper.

The rest of this paper unfolds as follows: in Section 2, we construct dollar-denominated well-being indices for nine countries using a mix of indicators that reflect economic, social, and developmental factors. Then, we normalize these indices, allowing them to be compared and aggregated into a global index that offers a holistic view of the global well-being. In Sections 3 and 4, we assess the stability and potential volatility of these indices using dynamic asset pricing, and we present the optimal portfolio composition and efficient frontier in Section 5. Section 6 introduces an option pricing model for the global socioeconomic well-being index, and this is followed by our concluding remarks in Section 7.

2. World Development Socioeconomic Well-Being Indicators

The World Bank (2022) compiles internationally comparable statistics on global development and poverty alleviation. This section outlines the world development indicators used to construct our USD-denominated socioeconomic index. Our index primarily incorporates indicators related to income, health, labor, education, the economy, and global connections. These include the Gini index for income distribution, total population, life expectancy at birth, and unemployment, which measure population health, labor, and education dynamics. The gross national income (GNI), consumer price index (CPI), and GDP measure income, savings, prices, the terms of trade, and the economic structure. Foreign direct investment reflects financial flows related to global connections. Further descriptions of these world development indicators (IBRD 2022) can be found in the online appendix. Whenever it is necessary, we compute missing data using multiple imputations with principal component analysis (Josse et al. 2011). We modify the world development indicators to construct the global dollar socioeconomic well-being index in Section 2.1.

2.1. Constructing a Global Dollar Socioeconomic Well-Being Index

In this subsection, we construct a global socioeconomic well-being index using the eight world development indicators (WDIs) defined by the World Bank (Gini index, total unemployment, life expectancy at birth, GNI per capita, consumer price index, population, foreign direct investment (in current USD), GDP/capita) for the nine countries in our study. Based on the data availability in IBRD (2022), we selected these nine countries from all the continents and extracted reported data from between 1990 and 2020.

We denote by $F(k,l)$ the k th world development indicator ($k = 1, \dots, K = 8$) for the l th country ($l = 1, \dots, L = 9$) for a given year such that all indicators are strictly positive; that is, $F(k,l) > 0$ for all k,l so that they positively contribute to the socioeconomic well-being index. Therefore, we transform the Gini index and unemployment indicators into the neg. Gini index (100-Gini index) and employment (100-unemployment), respectively. The US development indicators that positively contribute to the US socioeconomic well-being index are shown in Figure 1a.

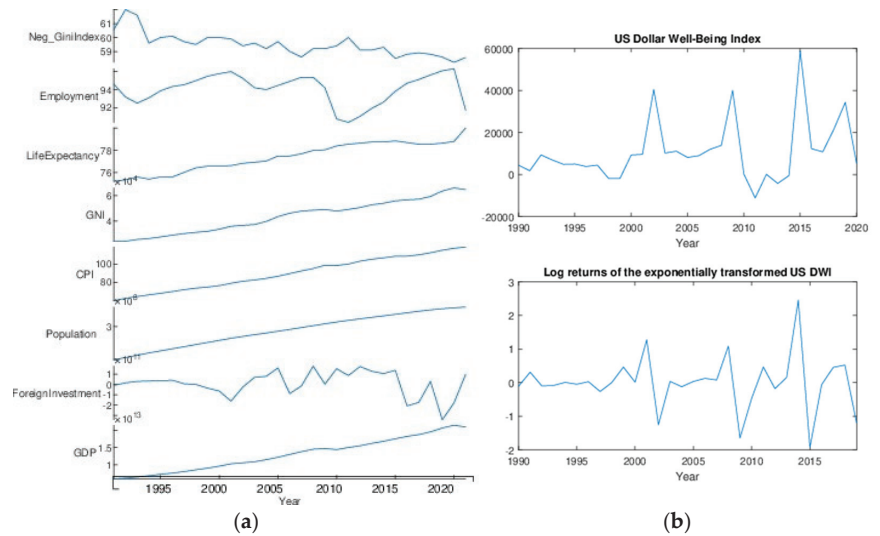


Figure 1. (a) US development indicators from World Bank reports (IBRD 2022) for 1990–2020 and (b) the US dollar socioeconomic well-being index (DWI) constructed using Equation (3) and the log returns of the exponentially transformed DWI with constraints from Equation (5).

In order to compare the significance of factors with respect to countries, we normalize each country's indicator, $F(k,l)$, using the corresponding indicator of all the countries ($F(k,l)$, $l = 1, \dots, L = 9$) as follows:

$$FN(k,l) = \frac{F(k,l)}{\sum_{l=1}^L F(k,l)}, \quad k = 1, \dots, K, \quad l = 1, \dots, L. \quad (1)$$

We then define the socioeconomic well-being index for country l , which we denote by $WI(l)$, as the average of its normalized indicators, excluding the GDP:

$$WI(l) = \frac{1}{K-1} \sum_{k=1}^{K-1} FN(k,l), \quad l = 1, \dots, L, \quad (2)$$

so that $WI(l) \in (0,1)$, $l = 1, \dots, L$. We monetize the US dollar value of $WI(l)$ at year t , $WI_t(l)$, by weighting it with its corresponding GDP/capita, $GDP_t(l)$, to define a US dollar-denominated index of socioeconomic well-being, the dollar socioeconomic well-being index (DWI), for country l at year t , as follows:

$$DWI_t(l) = GDP_t(l) \cdot WI_t(l), \quad t = t_0 = 1990, \dots, t = 2020. \quad (3)$$

This reflects the “well-being” of a resident in country l in US dollars. For instance, Figure 1b depicts the US per capita well-being. Figure 2a highlights variations in the well-being of different countries based on the dollar value per capita (DWI).

To compare the well-being of an individual with that of the overall population, we construct a global DWI by taking the average of the DWIs:

$$DWI_t = \frac{1}{L} \sum_{l=1}^L DWI_t(l), t = t_0, \dots, t_{30}. \quad (4)$$

The global DWI is shown in Figure 2b. The yearly deviations in terms of (thousands of) USD represent the change in the DWI per person from one year to another.

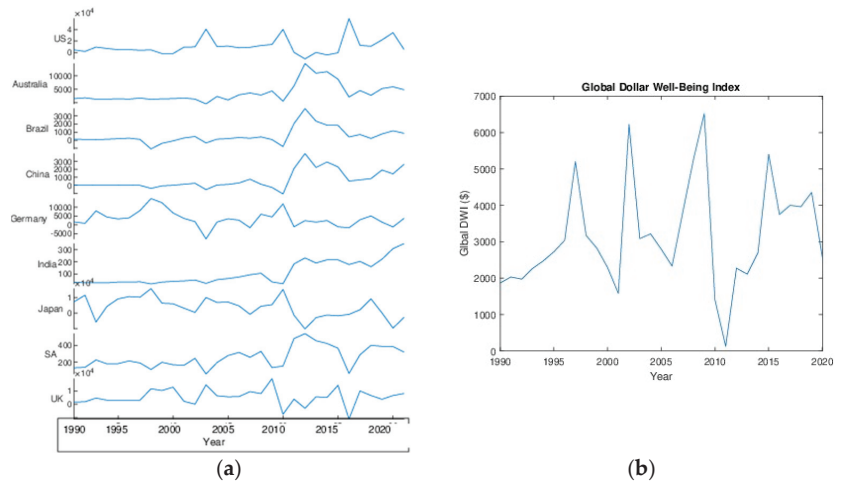


Figure 2. (a) Dollar socioeconomic well-being indices constructed using Equation (3) and (b) the global DWI proposed in Equation (4).

2.2. Econometric Financial Modeling of Socioeconomic Well-Being Indices

To model population well-being within a financial framework, we need to assign a US dollar value to the well-being of each resident in a given country, treating it as an asset price. This approach involves applying dynamic asset pricing theory to capture the financial market dynamics of these socioeconomic well-being indices over time. The exponential transformation of these time series is selected to mirror asset price dynamics, facilitating the application of dynamic asset pricing theory (Duffie 2010; Schoutens 2003).

We determine the exponential transformation for each DWI, considering the years $t = t_0, \dots, 2020$, as follows:

$$\begin{aligned} f(x) &= a \exp(bx), \quad a > 0, \quad b > 0, \\ f\left(\min_{l,t} DWI_t(l)\right) &= 0, \\ f\left(\max_{l,t} DWI_t(l)\right) &= 1. \end{aligned} \quad (5)$$

Under this optimization, we set the exponential transformation of the lowest DWI to 0 and that of the highest DWI to 1. We assume that the “asset price” (the happiness of the representative inhabitant) should have a minimal value of 0 and a maximal value of 1, which represents 100%. This scale is the one used in ESG rankings (Scatigna et al. 2021). The exponential transformation for the US DWI produces $a = 0.0037$ and $b = 0.0001$. With all DWIs being positive, we define the log returns for the exponentially transformed DWI and thus introduce the well-being asset pricing model for each country with respect to the global index.

We take the log returns of the exponentially transformed DWIs in the way that this is done in dynamic asset pricing theory (Duffie 2010):

$$R_t(l) = \log \frac{f(DWI_t(l))}{f(DWI_{t-1}(l))}; l = 1, \dots, 10, t = t_0, \dots, t_{30}, \quad (6)$$

where $f(DWI_t(l))$, $l = 1, \dots, 9$, is the exponentially transformed DWI for the l th country in year t , and we set $l = 10$ for the exponentially transformed global DWI. In each country, we model the log returns of the exponentially transformed DWI with an autoregressive AR(1) model:

$$R_t(l) = \phi_0 + z_t + \theta_1 z_{t-1}, t = t_0, \dots, t_{30}, \quad (7)$$

where $z_t = \sigma_t \epsilon_t$, and ϵ_t are assumed to be independent and identically distributed (iid) innovations, while ϕ_0 and θ_1 are new parameters to be estimated. We model the volatility (σ_t) using the best fit from among the time-varying volatility models ARCH(1), GARCH(1,1), and EGARCH(1,1). The GARCH(1,1) model is defined as

$$\sigma_t^2 = \alpha_0 + \alpha_1 z_{t-1}^2 + \beta_1 \sigma_{t-1}^2, t = t_1, \dots, t_{30}, \quad (8)$$

where ϕ_0 , θ_1 , α_0 , α_1 , and β_1 are parameters to be estimated (Bollerslev 1986). The sample innovations, ϵ_t , are iid random variables with zero mean and unit variance (Tsay 2005).

We model the log returns $R_t(l)$ in Equation (7) using the following univariate models with standard normal iid innovations (Hamilton 2020; Tsay 2005):

- Model 1: AR(1)-ARCH(1);
- Model 2: AR(1)-GARCH(1,1);
- Model 3: AR(1)-EGARCH(1,1).

We compare the performances of the three models based on the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), as shown in Table 1. For each country, we select the model that results in the lowest AIC and BIC among the estimated models (Models 1, 2, and 3). For example, Model 1 outperforms the other models in modeling the log returns of the US DWI. Combining the simulated sample innovations of each country, we get a 10-dimensional sample of 30 observations. We employ $S = 10,000$ scenarios for innovations based on a fitted 10-dimensional normal-inverse Gaussian (NIG) distribution, which is in the domain of attraction of a 10-dimensional multivariate Gaussian distribution (Øigård et al. 2004). By passing from normal to NIG innovations, we preserve the asymptotic unbiasedness of the parameters in the marginal time series models (Equation (8)). To assess the confidence bound for these parameters, we would need bootstrapping methods that are beyond the scope of this paper. Next, using the estimated parameters in Models 1, 2, and 3 for each of the 10 marginal (one-dimensional) time series, we generate $S = 10,000$ Monte Carlo scenarios of the innovations² for the log returns in the year 2021 ($(R_{t31}(1;s), \dots, R_{t31}(10;s)), s = 1, \dots, S$). As a result, we obtain S scenarios for the log returns of the 10 indices for the year 2021. As the innovations are models with an NIG distribution, which captures the tail dependencies of the indices, our overall forecast of the socioeconomic market for 2021 exhibits all the “stylized facts” of a financial market (see Taylor 2011; Cont 2001).

We use the simulated scenarios and the estimated parameters of the estimated univariate time series model (AR(1)-ARCH(1) or AR(1)-GARCH(1,1) or AR(1)-EGARCH(1,1)) to forecast S dynamic log returns for each DWI for 2021 (R_{2021}). For example, the estimated model for the US DWI log returns is an AR(1)-EGARCH(1,1) (with multivariate NIG sample innovations) model with the following parameters:

$$\begin{aligned} R_t(l) &= 0.09 + z_t - 0.13z_{t-1}, \\ \sigma_t &= \frac{z_t}{\epsilon_t}, \\ \sigma_t^2 &= -2.97 + 0.95z_{t-1}^2 - 0.93\sigma_{t-1}^2, t = t_0, \dots, t_{30}, \end{aligned} \quad (9)$$

where the ϵ_t are iid random normal innovations.

Table 1. Estimated dynamic model comparison for the log returns of DWIs in Equation (6) based on the AIC and BIC (Model 1: AR(1)-ARCH(1), Model 2: AR(1)-GARCH(1,1), Model 3: AR(1)-EGARCH(1,1)).

Country	Model 1	AIC Model 2	Model 3	Model 1	BIC Model 2	Model 3
US	2.4627	2.5209	2.5960	2.6495	2.7544	2.8498
Australia	1.5981	0.8630	0.5255	1.7849	1.0965	0.8058
Brazil	1.6695	1.7362	1.4492	1.8563	1.9697	1.7294
China	1.8168	1.4721	1.4353	1.0037	1.7056	1.7156
Germany	2.2366	2.3033	2.1957	2.4235	2.5368	2.4760
India	0.9228	0.9895	0.5342	1.1097	1.2230	0.8144
Japan	2.7870	2.8536	2.4123	2.9738	3.0872	2.6926
SA	2.1488	2.2155	2.1473	2.3357	2.4490	2.4276
UK	2.5711	2.6378	2.4302	2.7580	2.8713	2.7105
Global DWI	2.4346	2.5013	2.4220	2.6214	2.7348	2.7022

Then, we simulate S dynamic log returns for 2021, which we use as the dynamic asset prices for the US. In summary, we simulate dynamic log returns (for the year 2021, $t_{31} = 2021$) for each country by estimating econometric models for their historical log returns (between 2000 and 2020). The joint dependence of the log returns of all indices is determined by a multivariate NIG distribution on the sample innovations. Using the historical and dynamic econometric models of the indices as time series, we provide an asset valuation risk analysis. Note that historical time series analyses cannot be used for dynamic asset pricing, particularly for option pricing. Our econometric models are designed to be consistent with dynamic asset pricing theory, allowing the valuation of the indices as financial assets and the pricing of financial contracts, particularly insurance instruments, on the socioeconomic well-being indices.

3. Measuring the Tail Risk of Socioeconomic Well-Being Indices

In this section, we assess the DWI’s performance using economic factors linked to adverse changes in each country’s index, specifically evaluating its response to adverse events using the eight world development indicators. The Ljung–Box³ test indicates serial correlation and dependence in historical log returns. Our analysis emphasizes capturing linear and nonlinear return dependencies through dynamic log returns, which yield stationary time series. Initially, we fit multivariate NIG models to each country’s dynamic log returns (iid standardized innovations). Subsequently, we generate 10,000 scenarios from the multivariate NIG for scenario analysis and systemic risk measure computation.

To assess the tail risk, we calculate three systemic risk measures derived from the value at risk (VaR) (Philippe 2001). We calculate the VaR at the quantile levels of $(1 - \alpha)100\%$, with $\alpha = 0.05$ and $\alpha = 0.01$. For a given α -quantile level ($0 < \alpha < 1$), VaR_α is defined as follows:

$$VaR_\alpha(X) = -\inf\{x \in R | F_X(x) > \alpha\}, \quad 0 < \alpha < 1,$$
 (10)

where $F_X(x)$ is the cumulative distribution function of the log return X .

Tobias and Brunnermeier (2016) proposed a Δ -conditional value at risk (CoVaR) measure, which in our setting represents the log return of a DWI; it shows how the VaR of the financial system changes when a financial institution experiences distress relative to its median state. We use the conditional value at risk (CVaR) (Rockafellar and Uryasev 2000) to find the tail risk in the log returns of the DWI, denoted by X , at the α levels $(1 - \alpha)100\% = 95\%$ and $(1 - \alpha)100\% = 99\%$, as follows:

$$CVaR_\alpha(X) = \frac{1}{\alpha} \int_0^\alpha VaR_\gamma(X) d\gamma.$$
 (11)

If X has a probability density, then the CVaR coincides with the expected tail loss (ETL) or expected shortfall (ES) as follows (Rachev et al. 2008):

$$CVaR_{\alpha}(X) = -E(X|X \leq -VaR_{\alpha}(X)). \quad (12)$$

By switching from the VaR to the CVaR, the latter measure being a coherent risk measure,⁴ we may take into account episodes of more severe distress, back-test (Acerbi and Szekely 2014) the CVaR, and enhance its monotonicity concerning the dependence parameter. Girardi and Ergun (2013) improved the method for identifying whether a financial institution is considered to be in financial distress, from being exactly at its $VaR_{\alpha}(X)$ ($X = -VaR_{\alpha}(X)$) to being less than or equal to its $VaR_{\alpha}(X)$ ($X \leq -VaR_{\alpha}(X)$). Here, we refer to Y as the global DWI log returns and X as the log returns of the indices for each country in the alternative CoVaR defined in terms of the copulas in Mainik and Schaanning (2014). The CoVaR at the level of α , $CoVaR_{\alpha}$ (or ξ_{α}), is defined as

$$\xi_{\alpha} = CoVaR^{\alpha} = -F_{Y|X \leq F_X^{-1}(\alpha)}^{-1}(\alpha) = -VaR_{\alpha}(Y|X \leq -VaR_{\alpha}(X)), \quad (13)$$

where F_Y and F_X denote the cumulative distributions of Y and X , respectively, and $F_{Y|X}$ is the cumulative conditional distribution of Y given X . An extension of CoVaR, the conditional expected shortfall (CoES) for DWI log returns (Mainik and Schaanning 2014) at a level α , is given by

$$CoES_{\alpha} = -E(Y|Y \leq -\xi_{\alpha}, X \leq -VaR_{\alpha}(X)). \quad (14)$$

The conditional expected tail loss (CoETL) is the average of the DWI losses when the DWI and the country indices' extreme indicators are in distress (Biglova et al. 2014). The CoETL quantifies the portfolio downside risk in the presence of systemic risk. We define the CoETL at a level α as follows:

$$CoETL_{\alpha} = -E(Y|Y \leq -VaR_{\alpha}(Y), X \leq -VaR_{\alpha}(X)). \quad (15)$$

We assess the portfolio market risk by computing systemic risk measures (VaR, CoVaR, CoES, and CoETL) for the joint densities of the global DWI and individual country DWIs, as presented in Table 2. These values gauge the impact of a significant increase in each country's DWI on the global DWI. We choose confidence levels of 95% and 99% for statistical comparisons and disastrous loss calculations. The first column displays empirical correlation coefficients (Pearson's R), revealing no linear relationship between Germany and the DWI. Strong positive linear relationships exist between the DWI and the US and Japan. Conversely, China, Brazil, and South Africa (SA) exhibit robust negative relationships with the DWI, accompanied by low Pearson's R values that indicate weak relationships.

Columns 2 and 3 in Table 2 present each country's DWI VaR measures, which represent the maximum losses at the 95% and 99% confidence levels. These losses have 5% and 1% chances of surpassing the VaR threshold, respectively. Notably, the US experiences the maximum loss at the 99% confidence level for both dynamic and historical log returns. At the 95% confidence level, the highest loss is observed in the US (2.88) for dynamic data, while the UK records the highest loss (3.13) in historical data.

We measure the average loss beyond the 5% and 1% confidence levels using CVaR, which represents the expected loss in each country's worst 1% and 5% scenarios. Notably, the US registers the maximum average losses at both the 95% and 99% confidence levels. The CoES and CoETL assess how a significant decrease in each country's DWI affects the global DWI. For instance, at the 5% stress level, the CoES signifies the expected return on the DWI in the top 5% of each country's index. Remarkably, CoES5% and CoETL5% across all countries are quite similar, suggesting comparable impacts of a drastic DWI increase in each country on the global DWI. However, in the most severe scenario (1%), where the

stress on the US DWI has the greatest influence on the global DWI, Australia exhibits the lowest impact on the global DWI for dynamic log returns.

In summary, Table 2 presents the market risk based on left-tail systemic risk measures for the global DWI concerning each country’s DWI at varying stress levels (5% and 1%). These insights can aid investors in evaluating the global DWI market risk and optimizing their portfolios in the global financial market of socioeconomic well-being indices.

Table 2. Comparison of Pearson’s *R* and left-tail systemic risk measures (VaR, CoVaR, CoES, and CoETL) of the joint densities of the global DWI and DWI of each country at different stress levels for dynamic and historical log returns.

Dynamic Log Return Left-Tail Risk Measures									
Country	Pearson’s R	VaR %95	VaR %99	CVaR %95	CVaR %99	CoES %95	CoES %99	CoETL %95	CoETL %99
US	0.55	2.88	2.55	3.56	3.66	0.43	0.61	1.54	2.65
Australia	−0.25	0.52	0.39	0.52	0.66	0.47	0.65	0.31	0.53
Brazil	−0.63	0.96	0.52	0.70	1.24	0.47	0.65	0.59	0.96
China	−0.76	0.76	0.40	0.62	1.08	0.47	0.65	0.36	0.76
Germany	0.00	1.31	0.39	0.58	1.70	0.46	0.63	0.55	1.22
India	−0.41	2.11	1.10	1.39	2.89	0.47	0.65	1.06	2.12
Japan	0.40	0.60	0.26	0.47	0.75	0.45	0.63	0.28	0.58
SA	−0.58	1.53	0.84	1.11	2.00	0.47	0.65	0.89	1.54
UK	0.24	0.67	0.37	0.47	0.80	0.45	0.64	0.43	0.66
Historical log return left-tail risk measures									
US	0.79	2.74	4.87	4.30	5.66	1.85	3.83	1.16	4.70
Australia	−0.65	1.81	3.14	2.91	3.46	2.44	3.97	1.93	3.15
Brazil	−0.79	1.81	3.47	3.19	3.86	2.44	3.97	1.97	3.49
China	−0.82	2.09	3.73	3.36	4.23	2.44	3.97	2.18	3.75
Germany	0.04	2.35	2.80	2.67	2.97	2.44	3.97	2.35	2.80
India	−0.72	1.42	2.64	2.34	3.05	2.44	3.97	1.47	2.65
Japan	0.57	2.77	2.85	2.83	2.87	2.44	3.97	2.54	2.85
SA	−0.77	2.28	3.69	3.42	4.06	2.44	3.97	2.39	3.70
UK	−0.06	3.13	4.57	4.24	5.02	2.09	3.97	2.70	4.37

4. Regression and Jensen’s Alphas of the Socioeconomic Well-Being Indices with Respect to the Global Socioeconomic Well-Being Index

This section describes the estimation of the relationships between a country’s DWI and the global DWI. We use a regression analysis to capture the pairwise linear dependences between each DWI and the global DWI as follows:

$$Y_l = a_l + b_l Y + e_l, \tag{16}$$

where Y_l and Y are the log returns of the exponentially transformed DWI of country l and the global index, respectively. The terms a_l , b_l , and e_l denote the intercept, gradient, and random error corresponding to the regression line, respectively.

The ordinary least squares (OLS) method is typically used to estimate the parameters a_l and b_l when the errors follow a normal distribution. However, this assumption breaks down when there are outliers and highly influential observations. Robust regression (RR) addresses this issue by iteratively reweighting least squares to assign optimal weights to each data point (Hu et al. 2021; Knez and Ready 1997). OLS is suitable for short-term predictions, as it excludes outliers, while RR is preferred for longer-term forecasts.

We apply RR and OLS to the log returns of each DWI and the global DWI. For instance, we compare the regression analysis performance for the US DWI using data observed from 1990 to 2020 (historical regression) with the data generated for 2021 using time series modeling (dynamic regression), as discussed in Section 2.2. This involves regressing over $S = 10,000$ Monte Carlo scenarios for both individual and global indices.⁵ Table 3 presents estimated coefficients (intercept and gradient) for the regression models, along with goodness-of-fit measures (p -values, standard errors, and root mean square error

(RMSE)). In dynamic RR and OLS models, the estimated coefficients are quite similar to those in the historical models. However, dynamic regression models outperform historical ones, exhibiting lower *p*-values and standard errors. Moreover, dynamic RR surpasses dynamic OLS, with lower RMSEs and standard errors for the estimated coefficients.

Figure 3 shows the estimated regression lines for historical and dynamic data. The historical data, which are from 1990–2020, represent the average indicator values used in constructing the index, while the dynamic data comprise 10,000 generated scenarios for 2021. Historical regression is ineffective due to heavy-tailed asymmetry and volatility clustering in the data. Conversely, dynamic data conform to the normal distribution, as discussed in Section 2.2. The dynamic model captures autoregression, asymmetric volatilities, and heavy-tailed asymmetric copula dependence. Given increasing volatility over time, the dynamic model is preferable due to volatility clustering.

Table 3. Regression analysis for historical and dynamic log returns in the US.

		Regression Type			
Data Type	RR				
		Coefficient	p-Value	Standard Error	RMSE
Historical	Intercept (a)	−0.026	0.705	0.070	0.381
	Gradient (b)	0.673	0.000	0.093	
Dynamic	Intercept (a)	0.275	0.000	0.007	0.489
	Gradient (b)	1.313	0.000	0.017	
		OLS			
Historical	Intercept (a)	−0.008	0.937	0.093	0.512
	Gradient (b)	0.862	0.000	0.125	
Dynamic	Intercept (a)	0.378	0.000	0.008	0.603
	Gradient (b)	1.360	0.000	0.022	

Both historical and dynamic regressions for the US DWI yield upward forecasts (refer to Figure 3). Addressing the grim state of global well-being is essential. A steeper gradient in the dynamic regression signifies a more promising future well-being for a country. Table 4 presents estimated gradients from dynamic RR for all countries in the study. Among the considered countries, the US boasts the highest well-being, while South Africa ranks lowest. Detailed regression analyses for all countries are available from the authors.

Table 4. Robust regression for historical and dynamic log returns in the US. Regression lines for both historical and dynamic data result in upward forecasts.

Country	Estimated Gradient	Standard Error
US	1.36	0.0174
Japan	0.11	0.0025
UK	0.07	0.0031
Germany	0.01	0.0125
Australia	−0.22	0.0092
Brazil	−0.77	0.0095
China	−1.04	0.0087
India	−1.11	0.0025
SA	−1.14	0.0157

In the log return time series of the indices, negative drops are prominent, and these downturns are effectively captured by the NIG distribution of innovations (Schlösser 2011, pp. 129–63). Hence, we employ AR(1)-ARCH(1,1) or AR(1)-EGARCH(1,1) with a multivariate NIG distribution for option pricing, providing flexible probability distributions that account for heavy tails and dependencies, both centrally and in the tail. Dynamic models enable us to assess the value of insurance instruments accurately. Historical (static) methods fall short in providing suitable pricing models for socioeconomic well-being indices; they lack the ability to establish no-arbitrage values for insurance instruments.

Thus, we must rely on no-arbitrage asset pricing theory and implement dynamic models for pricing the well-being of individual countries and the global index. Therefore, we employ dynamic predictive models to forecast future trends in the global well-being that are country-specific and on a global scale.

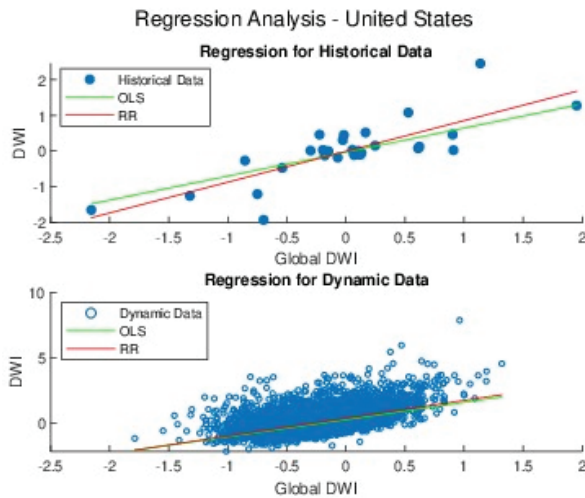


Figure 3. Robust regression for historical and dynamic log returns in the US. Regression lines for both historical and dynamic data result in upward forecasts.

Our global index (DWI) functions as a “market index”, while the US index serves as a “risky asset” within the socioeconomic well-being index market. Thus, we aim to determine the “beta” in the capital asset pricing model (CAPM) (Fama and French 2004) for the US index. If more countries are involved, we seek an analog of the Fama–French three-factor model (Fama and French 1992, 1993) for the socioeconomic well-being index market. Essentially, we want to identify which countries have the most significant impact on global well-being.

Jensen’s alpha (Jensen 1968) shows the average return on a portfolio or investment above or below the portfolio benchmark.⁶ We apply Jensen’s alpha to determine the maximum possible return on each country’s DWI and the performance of each country’s DWI compared to the global DWI.

Table 5 presents Jensen’s alpha values for each country’s DWI. The historical indices have nonsignificant alphas that are close to zero, indicating that there is no advantage compared to the global DWI. In contrast, the dynamic indices yield different results. South Africa and India show positive alphas (0.5359 and 0.5306, respectively), suggesting that their DWIs offer higher risk-adjusted returns. Japan, Germany, and the US exhibit statistically significant negative alphas, indicating underperformance compared to the global DWI. The UK, Australia, and China’s returns are roughly similar to the global DWI.

Table 5. Jensen’s alpha CAPM for historical and dynamic indices.

Country	Historical Index	Dynamic Index
Japan	−0.048	−0.5247
Germany	0.0087	−0.2615
US	−0.0075	−0.2264
UK	0.0232	0.1487
Australia	0.0242	0.189
China	0.0558	0.1951
Brazil	0.0198	0.4666
India	0.0894	0.5306
South Africa	0.0404	0.5359

5. Efficient Frontier of the Markets of Countries’ Well-Being Risk Measures

In accordance with Markowitz’s (1952) method, the goal of portfolio optimization is to identify the daily set of weights w that reduces the return risk of the portfolio (for that day) subject to a desired expected return (r_p). The targeted return value indicates the investor’s risk tolerance: the higher the value, the more risk the investor is willing to accept. Mean-variance and mean-CVaR optimization (Uryasev and Rockafellar 2001) have the goal of minimizing the portfolio variance σ_p and the portfolio CVaR, denoted by $\text{CVaR}_{p,\alpha}$, subject to a preferred expected return by using the variance and CVaR as the risk measure.

Consider a portfolio consisting of n risky assets with daily return values $r(t) = (r_1(t), r_2(t), \dots, r_n(t))$, with a portfolio mean and standard deviation $\bar{r} = (\bar{r}_1, \bar{r}_2, \dots, \bar{r}_n)$ and $(\sigma_1, \sigma_2, \dots, \sigma_n)$, respectively:

$$\text{minimize } w' \sigma_p w \text{ subject to } \bar{r} w = r_p \text{ and } \sum_{i=1}^n w_i = 1. \tag{17}$$

Because the target return varies, the best solution for (σ_p, r_p) results in a hyperbola curve known as the “efficient frontier” (EF), which is the region of the portfolio frontier where the projected mean returns exceed r_p . Consider a portfolio consisting of n risky assets with daily return values r_p , with the mean of the expected risk-adjusted returns denoted by $E(r_p)$ and the risk measure denoted by $V(r_p)$. The portfolio optimization can be summarized as follows:

$$\min_w (-\gamma E(r_p) - (1 - \gamma) V(r_p)) \text{ subject to } \sum_{i=1}^n w_i = 1, \tag{18}$$

where the risk-aversion parameter $\gamma \in [0, 1]$ determines the positions along the EF, with $\gamma = 0$ corresponding to the minimum-risk portfolio.

For each country’s DWI index, the optimization applied to the ensemble of the target returns produces an EF by considering the variance, $\text{CVaR}_{p,0.05}$, and $\text{CVaR}_{p,0.01}$ risk measures to contrast the effects of the central risk and tail risk on optimization and since the standard deviation is not a coherent risk measure. The DWI dynamic and historical indices referred to in Section 2.1 are used to illustrate the EFs of each country.

Figures 4 and 5a,b plot the EFs computed for the US dynamic and historical DWIs. The mean variance risk measure was used to generate efficient borders for the US dynamic and historical DWIs. We used the set of equally spaced values $\gamma = 0, 0.01, \dots, 0.99$ to plot each EF. Historical EFs are short since we have only 30 historical data points. We expect a long EF for simulated data. The standard deviation increased in the dynamic EF from 0.05 to 0.80 and in the same frontier for historical data from 0.08 to 0.34. The growth in $E(rp)$ for the dynamic and historical DWIs is consistent with the same observation; however, the increases are less evident for the EF.

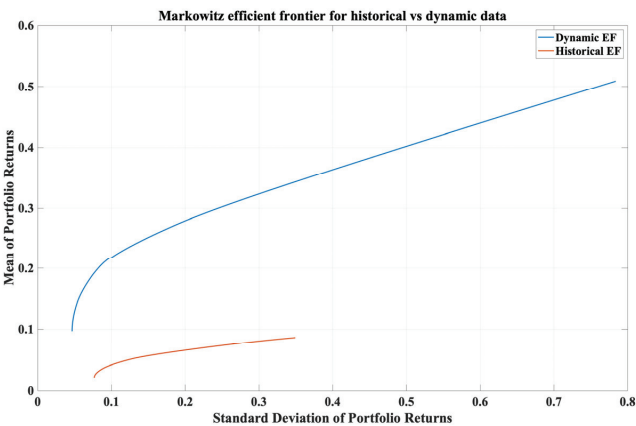


Figure 4. Markowitz efficient frontier.

Figure 6a,b reproduce Figure 4 for the tail risk measures $\text{CVaR}_{p,0.05}$ and $\text{CVaR}_{p,0.01}$. There are overall qualitative similarities in the behaviors of the CVaR EFs compared to the mean-variance EFs. While the EFs change “smoothly” and are convex, the variation in the behavior of the EFs is more pronounced under the CVaR risk measures compared to the mean-variance EF.

In Figures 7 and 8, we show the optimal portfolio weights on the EFs of the mean-variance and CVaR optimizations. For CVaR historical portfolio EFs, India and China’s weights increase with increasing standard deviation, emphasizing that these countries are at high risk, along with the US. Conversely, the dynamic portfolio EFs highlight that Brazil, China, and India show higher standard deviation levels. Additionally, the UK and South Africa become focal points at high CVaR risk levels, suggesting their safety in the event of a significant increase in the DWIs.

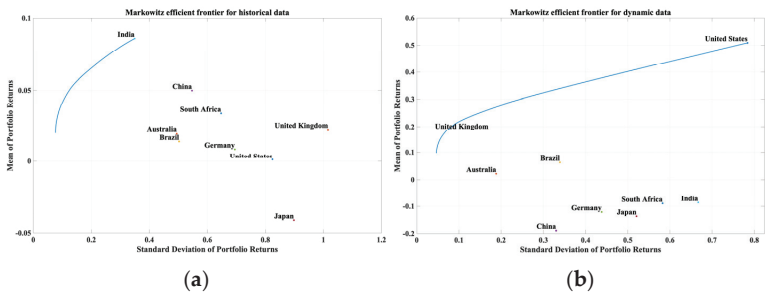


Figure 5. Markowitz efficient frontier: (a) historical indices and (b) dynamic indices.

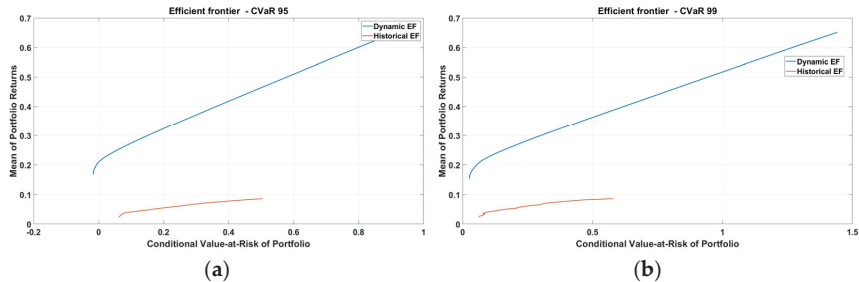


Figure 6. Conditional value-at-risk portfolio optimization: (a) $\text{CVaR}_{p,0.05}$ and (b) $\text{CVaR}_{p,0.01}$ EFs.

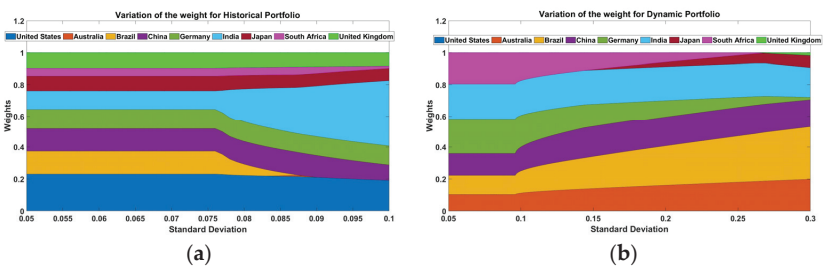


Figure 7. Variation of the weight composition of the Markowitz optimal portfolios along each efficient frontier (as a function of standard deviation): (a) historical portfolio and (b) dynamic portfolio.

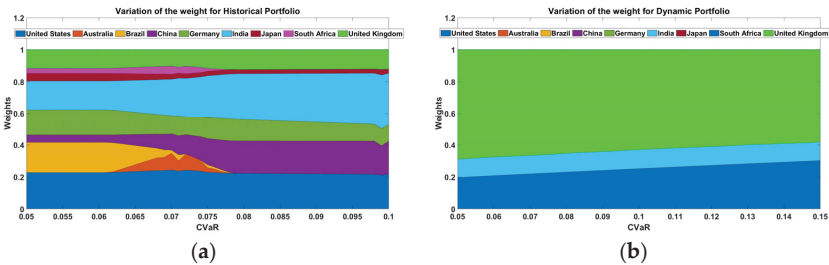


Figure 8. Variation of the weight composition of the $CVaR_{p,\alpha}$ optimal portfolios along each efficient frontier (as a function of α): (a) historical portfolio and (b) dynamic portfolio.

Efficient Frontier and Risk Measures of the Market for Countries with High GDPs

Economic rankings have evolved over time, and the nominal GDP/capita is a key metric for assessing national wealth. It is intriguing to examine the portfolio of socioeconomic well-being indices for the countries with the largest GDPs and compare them to the portfolios of all countries. As of 2022, the top four nations by nominal GDP were the US (USD 20.89 trillion), China (USD 14.72 trillion), Japan (USD 5.06 trillion), and Germany (USD 3.85 trillion), according to recent World Bank data. In this section, we analyze the portfolios of high-GDP countries; this is akin to the previous section focusing on the top four GDP nations. We employ optimization to create EFs, considering the variance, $CVaR_p$ at 0.05, and $CVaR_p$ at 0.01 risk measures to assess the effects of central and tail risk on optimization. We also introduce dynamic indices, as explained in Section 2.1, for these four countries, utilizing these new indices to illustrate their respective EFs.

Figure 9a,b display the EFs for the largest-GDP countries, comparing them to the EFs for all countries using mean-variance risk measures for both historical and dynamic DWI portfolios. The historical EF is relatively short due to limited historical data (30 data points), while the simulated data are expected to produce a longer EF. In the high-GDP countries' historical EF, the standard deviation increased from 0.16 to 0.55, and for all countries, it increased from 0.07 to 0.34. Interestingly, the growth in $E(r_p)$ for the DWI of all countries is higher than it is for high-GDP countries at the same risk level. This suggests that investing in high-GDP countries' DWI carries a higher risk for the same expected return compared to investing in the DWI from all countries. Qualitatively, the dynamic EFs for all countries and high-GDP countries exhibit similar behaviors, but the historical EFs for high-GDP countries display smoother, convex changes, while all countries' EFs show more pronounced fluctuations.

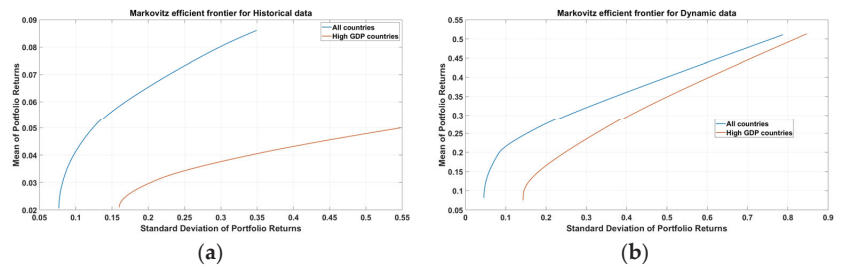


Figure 9. Markowitz efficient frontier for all countries and countries with high: (a) historical indices and (b) dynamic indices.

Figure 10a,b reproduce Figure 9 for the tail risk measures $\text{CVaR}_{p,0.05}$ and $\text{CVaR}_{p,0.01}$. The behaviors of the CVaR's EFs of the DWI for countries with high GDPs and those for all countries are comparable. Again, it can be seen that the risk of investing in the DWI for all countries is less than the risk for countries with high GDPs at the same expected return. The variation in the behavior of the EFs is more pronounced under the $\text{CVaR}_{p,0.01}$ risk measure.

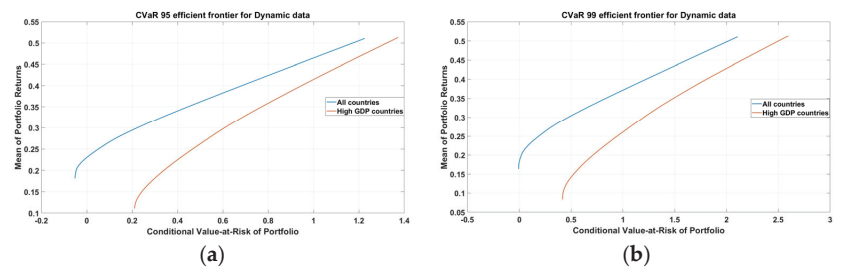


Figure 10. Comparing conditional value-at-risk portfolio optimization for dynamic indices: (a) $\text{CVaR}_{p,0.05}$ and (b) $\text{CVaR}_{p,0.01}$ efficient frontiers.

6. Pricing Options on Socioeconomic Well-Being Indices

In this section, we construct a financial model for pricing DWI options. Traditional methods like the Black–Scholes–Merton model, binomial option pricing, trinomial trees, Monte Carlo simulations, and finite difference models have been used to calculate option prices (Hull 2017; Duffie 2010). However, due to the presence of heteroskedasticity and heavy-tailed distributions in DWI returns, we avoid using the Black–Scholes–Merton model for DWI options. Instead, we rely on a discrete stochastic volatility-based model, specifically Duan's (1995) discrete-time GARCH approach with NIG innovations. This dynamic model provides an accurate pricing performance in a volatility-based framework. We provide DWI option prices, catering to institutional investors seeking to incorporate an additional socioeconomic dimension into their risk-adjusted portfolios and addressing potential mispricing events.⁷

In the standard GARCH(1,1) model, Blasild (1981) defines R_t for a given F_{t-1} , as distributed on a real-world probability space (P), as follows:

$$R_t \sim NIG\left(\lambda, \frac{\alpha}{\sqrt{a_t}}, \frac{\beta}{\sqrt{a_t}}, \delta\sqrt{a_t}, r'_t + m_t + \mu\sqrt{a_t}\right), m_t = \lambda_0\sqrt{a_t} - \frac{1}{2}a_t. \quad (19)$$

Gerber and Shiu (1994) describe the traditional approach to determining an equivalent martingale measure to obtain a constant option price. According to these authors, the Esscher transformation is the conventional method for deriving an equivalent martingale measure to obtain a constant option price. Since the moment-generating function of the

NIG distribution has an exponential form, the probability density of R_t is transformed into the risk-neutral probability density using the Esscher transform.

Chorro et al. (2012) discovered that R_t for a given F_{t-1} is distributed on the risk-neutral probability (Q) using the Esscher transformation as follows:

$$R_{-t} \sim NIG(\lambda, \frac{\alpha}{\sqrt{a_t}}, \frac{\beta}{\sqrt{a_t}} + \theta_t, \delta\sqrt{a_t}, r'_t + m_t + \mu\sqrt{a_t}), \quad (20)$$

where θ_t is the solution to $MGF(1 + \theta_t) = MGF(\theta_t) e^{r'_t}$, and MGF is the conditional moment-generating function of R_{t+1} given F_t .

Using Monte Carlo simulations, we construct future values of the DWI to price its call and put options as follows:

1. Fit GARCH(1,1) with NIG innovations to R_t and forecast a_1^2 by setting $t = 1$;
2. Repeat steps (a)–(d) for $t = 3, 4, \dots, T$, where T is the time to maturity of the DWI call option from $t = 2$:
 - (a) Estimate the model parameter θ_t using $MGF(1 + \theta_t) = MGF(\theta_t) e^{r'_t}$, where MGF is the conditional moment-generating function of R_{t+1} given F_t on P ;
 - (b) Find an equivalent distribution function for ϵ_t on Q ;
 - (c) Generate the value of ϵ_{t+1} under the assumption $\epsilon_t \sim NIG(\lambda, \alpha, \beta, \sqrt{a_t}\theta_t + \delta, \mu)$ on Q ;
 - (d) Compute the values of R_{t+1} and a_{t+1} using a GARCH(1,1) model with NIG innovations.
3. Generate future values of R_t for $t = 1, \dots, T$ on Q , where T is the time to maturity. Recursively, future values of the DWI are obtained as follows:

$$DWI_t = \exp(R_t) \cdot DWI_{t-1}. \quad (21)$$

4. Repeat steps 2 and 3 10,000 (N) times to simulate N paths to compute future values of the DWI.

Then, for a specific strike price K , the approximate future values of the DWI at time t are the average of the DWIs, and this price is used to determine the price of the call option ($\hat{C}$ and $\hat{P}$, respectively):

$$\hat{C}(t, T, K) = \frac{1}{N} e^{-r'_t(T-t)} \sum_{i=1}^N \max(DWI_T^{(i)} - K, 0), \quad (22)$$

$$\hat{P}(t, T, K) = (t, T, K) = \frac{1}{N} e^{-r'_t(T-t)} \sum_{i=1}^N \max(K - DWI_T^{(i)}, 0) \quad (23)$$

Call option pricing ($\hat{C}$) can help investors in planning to purchase our socioeconomic well-being indices at a predetermined strike price within a predetermined time frame (time to maturity).

Figure 11a shows call and option prices for the DWI, depending on the moneyness (S/K) and time to maturity (T). As the strike price rises, DWI call option prices slightly decrease, while an increase in the maturity time signifies a rise in DWI prices, reflecting increased population happiness within the financial market context. In Figure 11b, we display selling prices for index shares, explaining the relationship between put option prices ($\hat{P}$), the strike price (K), and the time to maturity (T). Put option prices are lower than call options with the same moneyness and maturity time, but they increase with higher strike prices, suggesting a potential linear relationship. The implied volatility, a reliable indicator of future volatility, is shown in Figure 12 for the DWI; it was constructed using the market values of call option contracts as proxies for upcoming event expectations. The implied volatility is determined based on the time to maturity (T) and moneyness ($M = S/K$). During intense market stress, an inverted volatility “grin” may appear on the observed volatility surface, with the highest implied volatilities corresponding to increasing moneyness.

Implied volatilities are higher for lower-strike-price options than for higher-strike-price options due to the downward-sloping volatility skew. As the maturity approaches 16 years, the implied volatilities tend to stabilize. These option prices are better suited for hedging than speculation, serving a role akin to that of portfolio insurance.

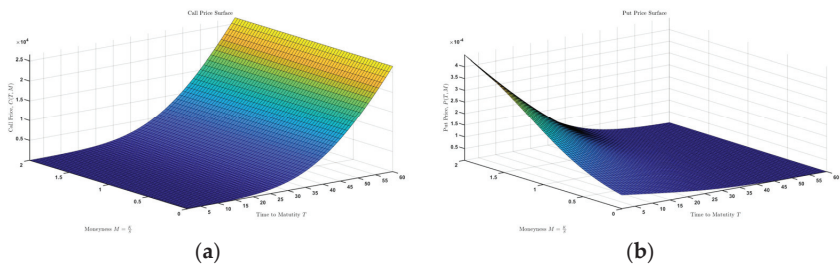


Figure 11. Option prices for the US DWI at time t with varying strike prices K using a GARCH(1,1) model with NIG innovations: (a) call prices and (b) put prices.

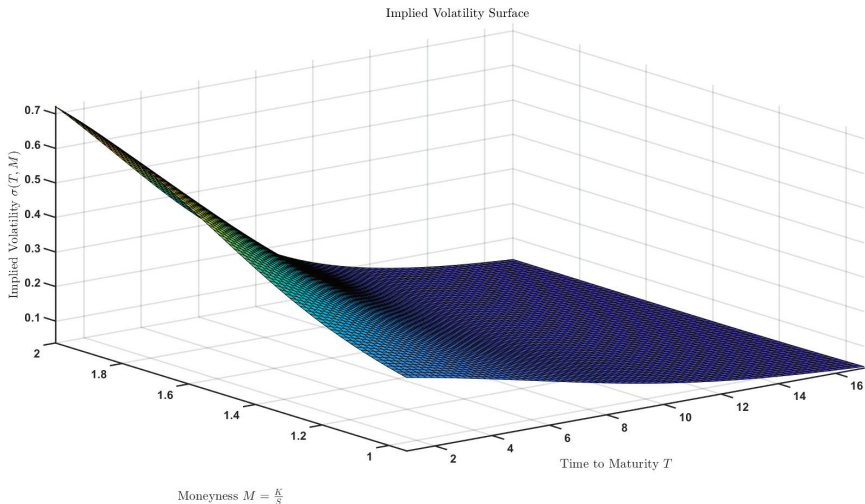


Figure 12. Implied volatilities of the US DWI based on the time to maturity (T) and moneyness ($M = S/K$) using a GARCH(1,1) model with NIG innovations.

7. Discussion and Conclusions

We introduced incorporating the well-being of the entire country into risk-return profiles for institutional investors. We then established a financial market for the socioeconomic well-being indices of nine countries, integrating eight world development indicators. This model enables the financial industry to assess, manage, and hedge against potential adverse index movements globally. We introduced a novel financial market framework for these indices, fostering active financial sector engagement in national well-being. Our approach involves risk assessment using dynamic asset pricing theory and the creation of hedging instruments, empowering institutional investors to trade these indices and establish insurance funds against adverse movements. This provides a new market for securities that protect against declines in social welfare, much like how traditional financial instruments protect against economic downturns.

While ESG financial markets address socioeconomic issues, our focus is broader, encompassing national citizens' well-being. We have offered option prices on these indices as insurance instruments against future index downturns. These findings help in estimating

the funds required to enhance global well-being, utilizing financial contracts for insurance against index declines. Developing an early-warning system for global socioeconomic index downturns requires additional data. Our indices are not tradable, but introducing exchange-traded funds (ETFs) would involve replicating their dynamics with fixed-income portfolios through asset liability management. For instance, an ETF replicating the US DWI could use put options as insurance instruments if the US DWI declined in 2024, ensuring that investors receive the put value on the ETF mirroring the US DWI.

Based on the data availability in IBRD (2022), we considered only using data from the US, Australia, Brazil, China, Germany, India, Japan, South Africa, and the UK as representative countries from all the continents, and extracted reported data from between 1990 and 2020. Since this provides a complete data set, missing data imputation is minimal, which minimizes the statistical error due to missing data imputation.

Our proposed method for constructing indices of socioeconomic well-being is not limited to the nine countries discussed in this paper. A similar portfolio analysis will be performed for various regions in the world based on geographical and economic criteria. This allows for addressing the following problems: which individual or group of countries contribute to the world's wellbeing? And to what degree? We will extend these financial management principles to construct a financial market of socioeconomic well-being for the organization of the petroleum exporting countries (OPEC) by developing the OPEC DWI. Then, we will check the contribution of each OPEC country to the OPEC DWI and determine what would be the effect if a country leaves or joins OPEC. We believe these findings will help the financial industry, which works with world organizations, to address the potential socioeconomic issues related to the well-being of these societies.

Author Contributions: Conceptualization, S.R.; methodology, F.J.F. and S.R.; software, T.V.M. and A.S.; formal analysis, T.V.M., A.S. and S.R.; investigation, T.V.M. and A.S.; data curation, T.V.M. and A.S.; writing—original draft preparation, T.V.M. and A.S.; writing—review and editing, T.V.M., A.S., F.J.F. and S.R.; visualization, T.V.M. and A.S.; supervision, S.R.; project administration, S.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset used in this study is publicly available from the World Bank, accessible through the following link: <https://datatopics.worldbank.org/world-development-indicators/> (accessed on 7 December 2023). All relevant data supporting our reported results can be accessed in this publicly archived dataset. We confirm our commitment to data transparency and appreciate the opportunity to provide this information.

Conflicts of Interest: The authors declare no conflicts of interest.

Notes

- ¹ The study, “Report by the Commission on the Measurement of Economic Performance and Social Progress”, was led by Joseph Stiglitz, Amartya Sen, and Jean-Paul Fitoussi (see Stiglitz et al. 2009).
- ² The Basel II Accord requires 10,000 scenarios in the generation of future portfolio returns to properly assess the tail risk portfolio of returns (Orgeldinger 2006; Jacobson et al. 2005).
- ³ For the Ljung-Box test, please refer to Ljung and Box (1978).
- ⁴ Regarding the CVaR as a coherent risk measure, please refer to Acerbi and Tasche (2002).
- ⁵ The historical regression assumes iid dependent variables. However, our econometric analysis shows that the dependent variables form a time series with characteristics quite different from white noise. The log returns of the socioeconomic well-being indices display heavy-tailed marginal distributions and volatility clustering. Thus, it is essential to employ a time series forecast, as demonstrated above, and conduct OLS and RR regressions on a sample of $S = 10,000$ Monte Carlo iid scenarios.
- ⁶ In real financial markets, Jensen's alpha is generally nonzero, as they often operate in a pre-equilibrium state with price fluctuations (Soros 2015).
- ⁷ For more information about discrete stochastic volatility-based models, refer to, for example, Duan (1995), Barone-Adesi et al. (2008), and Chorro et al. (2012).

References

- Acerbi, Carlo, and Balazs Szekely. 2014. *Backtesting Expected Shortfall: Introducing Three Model Independent, Non-Parametric Backtest Methodologies for Expected Shortfall*. Technical Report, Working Paper. New York: MSCI Inc.
- Acerbi, Carlo, and Dirk Tasche. 2002. Expected shortfall: A natural coherent alternative to value at risk. *Economic Notes* 31: 379–88. [CrossRef]
- Ahmad, Husnain F., and Irfan Qureshi. 2021. *Quantifying and Comparing Wellness across Nations: A Cross-Country Empirical Analysis*. Mandaluyong: Asian Development Bank. Available online: <https://www.semanticscholar.org/paper/Quantifying-and-Comparing-Wellness-Across-Nations%3A-AhmadQureshi/b02120cba7a744644b5532eebcfea6fc25ab7b00> (accessed on 7 December 2023).
- Barone-Adesi, Giovanni, Robert F. Engle, and Lorian Mancini. 2008. A GARCH option pricing model with filtered historical simulation. *Review of Financial Studies* 21: 1223–58. [CrossRef]
- Biglova, Almira, Sergio Ortobelli, and Frank Fabozzi. 2014. Portfolio selection in the presence of systemic risk. *Journal of Asset Management* 15: 285–99. [CrossRef]
- Blaesild, Preben. 1981. The two-dimensional hyperbolic distribution and related distributions with an application to Johansen's bean data. *Mathematical Finance* 68: 251–63.
- Bollerslev, Tim. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31: 307–27. [CrossRef]
- Chorro, Christophe, Dominique Gúegan, and Florian Ielpo. 2012. Option pricing for GARCH-type models with generalized hyperbolic innovations. *Quantitative Finance* 12: 1079–94. [CrossRef]
- Cont, Rama. 2001. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance* 1: 223–36. [CrossRef]
- Duan, Jin-Chuan. 1995. The GARCH option pricing model. *Mathematical Finance* 5: 13–32. [CrossRef]
- Duffie, Darrell. 2010. *Dynamic Asset Pricing Theory*. Princeton: Princeton University Press.
- Fama, Eugene F., and Kenneth R. French. 1992. The cross-section of expected stock returns. *Journal of Finance* 47: 427–65.
- Fama, Eugene F., and Kenneth R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33: 3–56. [CrossRef]
- Fama, Eugene F., and Kenneth R. French. 2004. The capital asset pricing model: Theory and evidence. *Journal of Economic Perspectives* 18: 25–46. [CrossRef]
- Gerber, Hans U., and Elias S. W. Shiu. 1994. Option pricing by Esscher transforms. *Transactions of the Society of Actuaries* 46: 99–191.
- Girardi, Giulio, and A. Tolga Ergun. 2013. Systemic risk measurement: Multivariate GARCH estimation of coVaR. *Journal of Banking & Finance* 37: 3169–80.
- Hamilton, James D. 2020. *Time Series Analysis*. Princeton: Princeton University Press.
- Hu, Yuan, W. Brent Lindquist, and Frank J. Fabozzi. 2021. Modeling price dynamics, optimal portfolios, and option valuation for cryptoassets. *Journal of Alternative Investments* 24: 75–93. [CrossRef]
- Hull, John C. 2017. *Options Futures and Other Derivatives*. Saddle Brook: Pearson.
- IBRD. 2022. World Development Indicators. Available online: <https://datatopics.worldbank.org/world-development-indicators/> (accessed on 7 December 2023).
- Jacobson, Tor, Jesper Lindé, and Kasper Roszbach. 2005. Credit risk versus capital requirements under Basel II: Are SME loans and retail credit really different? *Journal of Financial Services Research* 28: 43–75. [CrossRef]
- Jensen, Michael C. 1968. The performance of mutual funds in the period 1945–1964. *The Journal of Finance* 23: 389–416.
- Josse, Julie, Jérôme Pagés, and François Husson. 2011. Multiple imputation in principal component analysis. *Advances in Data Analysis and Classification* 5: 231–46. [CrossRef]
- Knez, Peter J., and Mark J. Ready. 1997. On the robustness of size and book-to-market in cross-sectional regressions. *Journal of Finance* 52: 1355–82.
- Ljung, Greta M., and George E. Box. 1978. On a measure of lack of fit in time series models. *Biometrika* 65: 297–303. [CrossRef]
- Mainik, Georg, and Eric Schaanning. 2014. On dependence consistency of CoVaR and some other systemic risk measures. *Statistics & Risk Modeling* 31: 49–77.
- Markowitz, Harry. 1952. Portfolio selection. *Journal of Finance* 7: 77–91.
- Matos, Pedro. 2020. *ESG and Responsible Institutional Investing around the World: A Critical Review*. Charlottesville: CFA Institute Research Foundation.
- OECD. 2023. OECD Better Life Index. Available online: <https://www.oecdbetterlifeindex.org> (accessed on 7 December 2023).
- Orgeldinger, Jörg. 2006. Basel II and project finance: The development of a Basel II conforming rating model. *Journal of Structured Finance* 11: 84–95. [CrossRef]
- Øigård, Tor A., Alfred Hanssen, and Roy E. Hansen. 2004. The multivariate normal inverse Gaussian distribution: EM-estimation and analysis of synthetic aperture sonar data. Paper presented at 2004 12th European Signal Processing Conference, Vienna, Austria, September 6–10; pp. 1433–36.
- Philippe, Jorion. 2001. *Value at Risk: The New Benchmark for Managing Financial Risk*. New York: McGraw-Hill.
- Quiros-Romero, Gabriel, and Marshall B. Reinsdorf. 2020. *Measuring Economic Welfare: What and How?* Policy Paper No. 2020/028. Washington, DC: International Monetary Fund.
- Rachev, Svetlozar T., Frank J. Fabozzi, and Stoyan V. Stoyanov. 2008. *Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization: The Ideal Risk, Uncertainty, and Performance Measures*. Hoboken: John Wiley & Sons.
- Rockafellar, Ralph T., and Stanislav Uryasev. 2000. Optimization of conditional value-at-risk. *Journal of Risk* 2: 21–42. [CrossRef]

- Sagar, Ambuj D., and Adil Najam. 1998. The human development index: A critical review. *Ecological Economics* 25: 249–64. [CrossRef]
- Scatigna, Michela, Dora Xia, Anna Zabai, and Omar Zulaica. 2021. Achievements and Challenges in ESG Markets. Bank for International Settlements. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3989761 (accessed on 7 December 2023).
- Schlösser, Anna. 2011. *Pricing and Risk Management of Synthetic CDOs*. New York: Springer.
- Schoutens, Wim. 2003. *Lévy Processes in Finance: Pricing Financial Derivatives*. Hoboken: John Wiley & Sons.
- Soros, George. 2015. *The Alchemy of Finance*. Hoboken: John Wiley & Sons.
- Stiglitz, Joseph E., Amartya Sen, and Jean-Paul Fitoussi. 2009. *Report by the Commission on the Measurement of Economic Performance and Social Progress*. Paris: Commission on the Measurement of Economic Performance and Social Progress.
- Taylor, Stephen J. 2011. *Asset Price Dynamics, Volatility, and Prediction*. Princeton: Princeton University Press.
- Tobias, Adrian, and Markus K. Brunnermeier. 2016. CoVaR. *American Economic Review* 106: 1705–41.
- Trindade, A. Alexandre, Abootaleb Shirvani, and Xiaohan Ma. 2020. A socioeconomic wellbeing index. *arXiv*, arXiv:2001.01036.
- Tsay, Ruey S. 2005. *Analysis of Financial Time Series*. Hoboken: John Wiley & Sons.
- Uryasev, Stanislav, and Ralph T. Rockafellar. 2001. Conditional value-at-risk: Optimization approach. In *Stochastic Optimization: Algorithms and Applications*. Boston: Springer, pp. 411–35.
- World Bank. 2022. World Bank Group. Available online: <https://www.worldbank.org/en/home> (accessed on 7 December 2023).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



What's Wrong with Enterprise Risk Management?

John Fraser ^{1,†}, Rob Quail ² and Betty Simkins ^{3,*}

¹ Hydro One Networks Inc., Toronto, ON M5G 2P5, Canada

² Robert Quail Consulting, Toronto, ON M4E 3J5, Canada; quail.rob@gmail.com

³ Department of Finance, Spears School of Business, Oklahoma State University, Stillwater, OK 74078, USA

* Correspondence: betty.simkins@okstate.edu; Tel.: +1-405-744-8625

† Deceased.

Abstract: Enterprise risk management (ERM) was introduced in the 1990s and has since become expected by boards of directors and regulators as a sign of good management and good corporate governance. However, many organizations struggle to implement ERM, and still seek practical advice on ERM implementation. This article explains many of the reasons why organizations are unsuccessful in their efforts at implementation and provides practical solutions provided by an experienced risk manager and consultant, an ex-Chief Risk Officer, and an academic, all of whom have written extensively on the subject. This article should be of interest to practitioners involved in implementing ERM, to consultants in ERM, and to academics teaching courses on ERM, risk management, and related topics. This article also provides a base against which further future research can be performed as ERM practices continue to evolve.

Keywords: enterprise risk management; corporate governance; risk; risk management; COVID-19

1. Introduction

Over the last twenty years, enterprise risk management (ERM) has gained greater recognition as a key factor in good management and good corporate governance. Many regulatory bodies are now requiring boards of directors to be more involved in overseeing risk management, and most companies are seeking to implement ERM or key aspects thereof. For example, the Dodd–Frank Wall Street Reform and Consumer Protection Act recommended that foreign bank holding companies and U.S. bank holding companies (those over a certain size) establish a risk committee of the board, among other risk-related requirements.

Unfortunately, the success of ERM implementation has been less than expected. Beasley and Branson (2022) found that only 56% of large corporations reported having complete ERM processes in place. We have found that many organizations struggle with the implementation of ERM despite copious articles and the use of consultants.

In this *Journal of Risk and Financial Management* (JRFM) Communication article, we strive to fill an important gap in the literature. In the next section, we provide a brief background and review of the literature. In Section 3, we explain the reasons for these failures and provide practical advice on how to succeed. Based on our observations and experiences, we identify seven common problems with the way organizations go about implementing ERM. In the final section, we conclude, and describe future directions for ERM.

2. A Brief Background and Review of the ERM Literature

ERM can be viewed as the natural evolution of risk management. Kloman and Fraser (2021) put this development in context by providing a brief history of risk management, beginning in antiquity and leading up to 2019, including the description of several ERM milestones. In 1993, James Lam became the first ever ‘Chief Risk Officer’ (CRO) and is

Citation: Fraser, John, Rob Quail, and Betty Simkins. 2024. What's Wrong with Enterprise Risk Management? *Journal of Risk and Financial Management* 17: 274. <https://doi.org/10.3390/jrfm17070274>

Academic Editor: Thanasis Stengos

Received: 10 November 2023

Revised: 15 June 2024

Accepted: 19 June 2024

Published: 29 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

widely recognized as a pioneer in the field of ERM. Because listed firms and organizations must employ an independent senior executive with responsibility for the risk management function (commonly referred to as the CRO), the CRO position has become common.

Many other articles and books have been written about ERM, including academic research, case studies, 'how-to' books/chapters, surveys, and white papers, among other formats. Regarding academic research, Pagach and Pascanik (2021) provide an excellent review of the literature and focus on four specific areas of ERM research: (1) academic studies investigating firms adopting ERM, (2) studies that identify firms' characteristics associated with ERM implementation, (3) research aimed at determining whether ERM creates value for organizations, and (4) academic business case studies on ERM. Refer to Exhibit 39.2 of their article for the 20 most-cited ERM academic journal articles.

Several case studies of leading organizations have been published referencing the successful implementation of ERM. For a book containing many ERM case studies on ERM best practices, refer to (Fraser et al. 2015). One of the earliest case studies is about the ERM early adopter, energy giant Hydro One, Inc. (Aabo et al. 2005; Mikes 2008; Fraser et al. 2021c). Hydro One has been at the forefront of ERM for many years, especially in utilizing a holistic approach to managing risks, and serves as a best-practice case study for other firms to follow.

Narvaez (2011) includes case-study examples to explain how various public organizations have adopted ERM. She provides examples from organizations such as the Department of Homeland Security, the Centers for Disease Control and Prevention, the Department of Education, the University of California, and Dallas Fort Worth International Airport.

Other examples of ERM best-practices case studies include the following:

- Mars, Inc. (Warner 2015): Mars is a global food company and one of the largest privately held corporations in the U.S. This case study describes how Mars recognized the importance of providing its managers with a tool to take risks knowledgeably and comfortably in order to achieve its long-term goals.
- Statoil (Alviniussen and Jankensgård 2015): In this case study, the authors discuss ERM at Statoil, one of the top oil and gas companies in the world, now named Equinor. The case describes how, at Statoil, understanding and managing risk is considered a core value of the company, one which is written into the corporate directives and widely communicated to employees. ERM is thoroughly embedded in the organization's work processes, and its risk committee has managed the transition from a 'silo' mentality to the promotion of Statoil's best interests.
- University of California Health System (Crickette 2015): Crickette describes ERM at the University of California's (UC) Health System (composed of numerous clinical operations, including five medical centers). She describes how ERM plays an important role at the UC Health System and assists the organization in assessing and responding to all risks (operational, clinical, business, accreditation, and regulatory) that affect the achievement of the strategic and financial objectives of the organization.
- Bank of Tokyo-Mitsubishi (Nagumo 2005): Nagumo describes how the international banking giant Bank of Tokyo-Mitsubishi launched a global balanced scorecard as an enterprise-wide strategic management tool and integrated it with ERM.

There are many books providing advice on ERM. A few examples are described below:

- *Enterprise Risk Management: From Incentives to Controls*, 2nd Edition (Lam 2014): This book focuses on the 'what' of ERM. Lam describes both the art and science of effective ERM practices. This book covers key concepts, processes, and tools underlying risk management and highlights strategies to manage risk.
- *Implementing Enterprise Risk Management: From Methods to Applications* (Lam 2017): This book focuses on the 'how' of ERM implementation.
- *Enterprise Risk Management: Today's Leading Research and Best Practices for Tomorrow's Executives*, 2nd Edition (Fraser et al. 2021c): This book provides in-depth insights into what ERM managers are doing. This book also includes a chapter on teaching ERM,

and the pedagogical techniques described are being used at universities in the U.S. and Europe (see Lange and Simkins 2021).

Despite the wealth of published studies on the subject, many organizations still struggle with ERM implementation. Clearly ERM is not a one-size-fits-all exercise. Our article strives to fill this gap in the literature and hopefully will assist more organizations in achieving success in ERM implementation. In the next section, we describe common problems with ERM implementation and offer suggestions to overcome these challenges.

3. Common Problems with ERM Implementation

In this section, we describe in detail seven common problems that organizations encounter, impeding their ERM implementation. We provide guidance on how to avoid these pitfalls so that an organization can excel at ERM.

These common problems are the following:

1. An over-emphasis on reporting;
2. Not enough injection into the decision-making processes;
3. Too much adherence to a static process;
4. Treating risks as discrete items;
5. The misuse of models;
6. The belief that all risk is bad;
7. A lack of role-clarity.

3.1. An Over-Emphasis on Reporting

Here is a brief summary of what ERM was initially intended to accomplish. In 2000, before the International Organization for Standardization (ISO) and Committee of Sponsoring Organizations of the Treadway Commission (COSO 2017) standards were published, many early adopters of ERM sought to improve their competitive performance and increase value through rational, strategic, coordinated risk-taking; they sought to build transparent cultures where people were comfortable openly discussing risks and sharing their honest opinions about ways they might fail, i.e., to stimulate the conversations about uncertainty which are used to prioritize actions or alternatives or to allocate resources. From time to time, these conversations would result in artifacts such as reports or lists of risks and associated treatment strategies, which might be shared periodically with boards of directors and oversight groups to stimulate more conversations and build understanding of the role of uncertainty in the business. Reporting was a by-product of ERM, and not its central purpose.

Several things then happened which knocked many ERM practitioners off of this course in the early days. The first of these was the Sarbanes-Oxley Act (SOX) in 2002 (Sarbanes-Oxley Act 2002). At the time, many organizations ‘lumped SOX in’ with ERM. ERM conference agendas suddenly shifted from being about building more effective tools and processes to understand risk, to becoming focused on SOX implementation and reporting. In this way, ERM became seen as being more about assurance as to current controls, and about the present state of compliance, rather than uncertainties about the future.

The second disruptor was the widespread introduction of Governance, Risk, and Compliance (GRC) systems. This resulted in the emergence of increasingly intrusive and complex reporting tools, more and more elaborate reporting systems requiring many people to enter data into large databases for the purpose of tracking and reporting on compliance.

And then, third, there was a growing demand from boards for more and more risk reporting. On the one hand, that is a good thing: boards are becoming more sophisticated in terms of their understanding of their role in risk oversight (Fraser 2016). However, the problem is that, for many organizations, the central role of the ERM group has become the production of lists and reports on risks intended for consumption by the board, as opposed to actually helping the organization by enabling better decision making under uncertainty. One board director at a large, fast-growing company recently complained to one of the authors that the only time he ever saw risk assessments associated with major

proposals was post hoc, i.e., after the choice had been made. There were no conversations; it was purely one-way presentations to justify recommended courses of action. In his view, the organization was missing out on the benefits of discussing uncertainties associated with these major strategic choices with the board and receiving their guidance, advice, and perspective.

As a consequence of this massive shift toward ERM-as-reporting, in many organizations, managers' and executives' only interactions with their ERM groups consist of feeding periodic, quarterly, or annual bottom-up reporting cycles. ERM is not seen by them as a tool for making better decisions. Instead, it is seen as a way to describe or justify choices already made, or to placate senior management, boards, or oversight groups.

Recent experiences with COVID-19 are another good illustration of this. One of the authors is a member of the Strategic Risk Council (SRC) for the Conference Board of Canada, a round table of Chief Risk Officers (CROs) and other senior risk executives from a variety of sectors. When the COVID-19 pandemic first emerged, it seemed like an excellent research opportunity. Businesses were going through a once-in-a-century crisis, and all were going through it together, so it seemed like a good idea to diarise the collective experience of all these organizations' risk management groups. So, the SRC enlisted many CROs and had three or four different small-group video conference calls each week in which they shared what they were dealing with in real time. What they learned from those sessions was very disappointing, because many were working on similar, short-term problems such as how to have people working remotely, how to procure and distribute sufficient personal protective equipment, how to maintain supply chain integrity, and other immediate matters. Many CROs were engaged in assembling packages to placate their boards of directors and convince them that they were prudently managing the risks associated with the crisis. But few executives seemed to be asking the important risk-related questions, such as "What are the long-term implications for our business of this global event?" or "How might this event affect markets, or our competitive position with our customers or stakeholders?" Few seemed concerned with exploring the strategic implications of the pandemic. One organization actually shut down their entire ERM program for two years so the staff could help manage the current crisis.

The global pandemic presented a wonderful opportunity to demonstrate the value of ERM, and to help figure out these issues and understand the risks and opportunities. Surprisingly, most risk managers missed it. Why did they miss it? In many organizations, the ERM group is not seen as a support resource for making better decisions under uncertainty. Instead, they are seen as the 'heat-map people'.

Similarly, partway through the global pandemic, we published a paper providing recommendations as to how organizations should seize the opportunity during the COVID-19 crisis to consider the longer-term implications of risks to their operating environments (see Fraser et al. 2021a). This advice was largely ignored.

3.2. Not Enough Injection into Decision-Making Processes

In the 1980s, Nike wanted to 'own' the basketball shoes business. So, they went into big cities in the U.S. where kids were playing streetball. And they found the best players and gave them Nike shoes. They gave their shoes to the 'cool kids' so the other kids could see the 'cool kids' use the shoes as a tool to succeed, and so the other kids would want them, too (see, for example, Stonebrook 2021).

This is an excellent metaphor for the promotion of ERM within organizations. Find the people, either with the powerful processes or working on the messiest problems, who are open minded and willing to consider innovative approaches, and 'give them sneakers', i.e., give them tools, help support them in making better decisions and also in communicating the rationales for their decisions to their bosses. Almost without exception, in the authors' experience, every organization that has had ERM as a leading practice for a decade or more has followed that kind of approach: they identified the important processes or problems

in their business, and then found ways to customize their ERM approach so that it was a provocative force for improving those processes and helping to enable better decisions.

ERM must not be a silo. It must be a provocative agent, a disruptor. It is most effective when it gets injected into the most powerful, complicated, and risky decision-driven processes in the enterprise.

The following are several examples of specific ways to improve the impact of ERM through this kind of process-injection:

- Strategic planning. Risks are represented in the O (opportunities) and the T (threats) of SWOT. It is a natural fit.
- Business planning. Using bottom-up risk assessments as a driver for decision making and trade-offs about where to allocate resources, and ensuring the plans are aligned with the business strategy.
- Outsourcing. An outsourcing contract is not just the allocation of money and services; it also allocates risks between the two parties. So, why not transparently identify the risks related to the body of work that is being outsourced, use the contract to explicitly assign the risks between the two parties, and figure out how to make sure that they are oriented towards managing the critical things? (Quail 2021c).
- Investment prioritization. For every dollar available to the business, where should that money be used? One of the factors should pose the question of where that dollar can do the best in terms of mitigating risks and managing uncertainty about the achievement of the business objectives. Embed risk management right into those prioritization processes (Toneguzzo 2021).
- Technology projects. There is a saying that there is no better way for a Chief Information Officer to lose their job than to try and replace the enterprise resource platform or other enterprise technology in the business. It is risky work. Therefore, it follows that risk assessments should be done; not just according to the Project Management Office tools and methods, but in terms of scoping, resourcing, timing, vendor selection, and the all-critical go/no-go decision before 'go-live' and the resulting effects (Winters 2021).
- Regulatory compliance management. Many businesses are involved in very complicated regulatory environments, and it can be a challenge for organizations to prioritize areas for the allocation of resources for compliance management and control. Risk management can help organizations prioritize resources by exploring this question: Which of these regulatory requirements, if not met, has the bigger potential to cause harm or affect the achievement of the stated business objectives?

3.3. Too Much Adherence to a Static Process

The ISO 31000 risk management process is a good one (see ISO 2018). But one of the problems with any standardized process is that it implies that there is only one way to go about it. There are some great thinkers on the topic of ERM in academic circles, in particular Anette Mikes, Associate Professor of Accounting, Saïd Business School, University of Oxford, and Bob Kaplan, Senior Fellow and Marvin Bower Professor of Leadership Development, Emeritus at the Harvard Business School, who have written that they believe that the ISO and COSO standards came out too early, and that these standards have been a drag on innovation ever since. The models suggest that there is but one process that is the pathway to 'goodness' in terms of managing risks as a business (Kaplan and Mikes 2012).

But there are other useful, valid risk management processes; here are some examples:

- Black Swans, or extreme-end-of-tail risks (Taleb 2010). These risks are very unlikely but have the potential for extreme impact. The tools that one normally uses for prioritizing risks do not work anymore. Instead, one needs to identify other ways to learn from potential Black Swan-type scenarios as a kind of thought-experiment, i.e., if one of these scenarios occurs, is the organization resilient enough to be able to react in time, or at least faster and better than the competitors?

- Scenario planning exercises as pioneered by Royal Dutch Shell (see Schwartz 1991; Wilkinson and Kupers 2013). Remember, the ISO definition of risk is the effect of uncertainty on objectives. What scenario planning does is test to see whether those are the right objectives in the first place. In this way, it is about risks not to the strategy, but of the strategy. We note that, in the wake of COVID-19 and climate change concerns and the war in Ukraine, there has been a recent resurgence in the popularity of scenario planning.
- Custom criteria should be developed to help inform decision making, e.g., developing a technology road map for an organization by applying things like priority and opportunity in capacity, as well as an assessment of risks in the organization’s ability to deliver.
- ERM can be dovetailed into strategy-setting through exercises like the risk appetite process (Quail 2021b; Ismail 2021).

In other words, there are ways to assess uncertainty and prioritize risks without going through the same old identify–assess–evaluate process as described by the ISO standard.

3.4. Treating Risks as Discrete Items

Most risk professionals who work in ERM generate lists of the top risks. Risks are brainstormed or extracted from a risk universe, or from a registry or some other source. They are assessed individually using some set of criteria and ranked. There may be some kind of roll-up for the organization or perhaps workshops or risk assessments with the executive team and senior levels of management. By whatever means, the risks assessed are ranked and delivered to the folks in charge to improve their understanding, and hopefully to stimulate conversations. For more information on risk management workshops, refer to (Quail 2021a).

However, risks do not act in isolation. Risks are not discrete elements. They are interconnected networks. ERM is a dynamic ecosystem. Failure of one part of the business can affect the success of others. For example, surprises in human resources might affect the performance in technology, and vice versa. Not enough work tries to understand and depict how risks interact or to fully map out the implications if a risk event actually occurs in one area.

Figure 1 provides a rudimentary example of a visual tool that can be applied to understand how risks are interconnected. In this fictitious organization that has outsourced its I.T. infrastructure management to control its risk and reduce cost, the risk of mismanagement of outsourcing can affect other things on the map. In the figure, the downstream impact of this outsourcing governance failure on other risks is depicted graphically. This is one simple example provided to illustrate what every risk manager should be doing—this kind of modeling, and the sharing of it with decision makers.

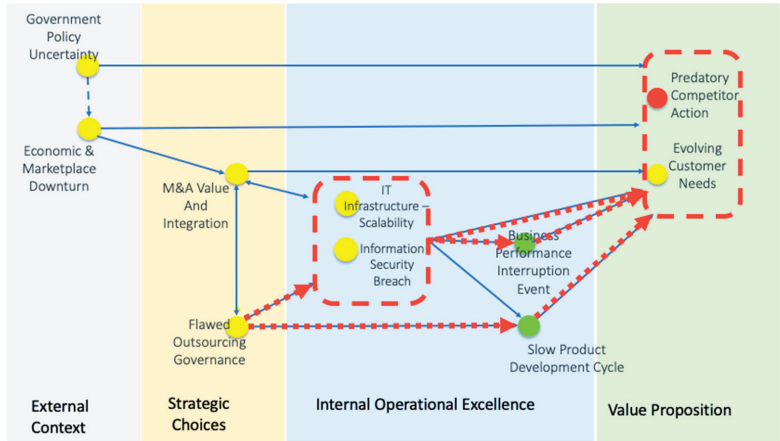


Figure 1. Graphical representation of risks’ interconnectedness.

In the wake of the COVID crisis, which was immediately followed by the war in the Ukraine, businesses have been operating in a crisis mode or dealing with various fragilities or discontinuities in their business environment. Many aspects of their business, such as the supply chain or human capital, remain fragile. One should fully expect that any given organization will experience at least one major surprise—a ‘second punch’—in the next few years. The question is the following one: can they take that punch, and do they know how risks could materialize, multiply, and/or accumulate? For additional information, see Winters (2021).

3.5. The Misuse of Models

Models are useful, but they are almost always wrong. In 1976, a British statistician named George Box wrote the famous line, “All models are wrong, some are useful.” The real world is much more complicated a proposition than we can ever represent in a single model.

Figure 2 was posted on LinkedIn in 2021 by a ‘risk expert’ as a way to try to convince other people to use their services. What they said was that there are two potential investments, namely, Investment A at the top, and Investment B at the bottom, and there is a heat-map for each. Readers were supposed to choose one of the two investments based on the heat-map. The question itself was alarming in its naiveté. What was even more alarming, however, was the number of other ‘experts’ who tried to answer it.

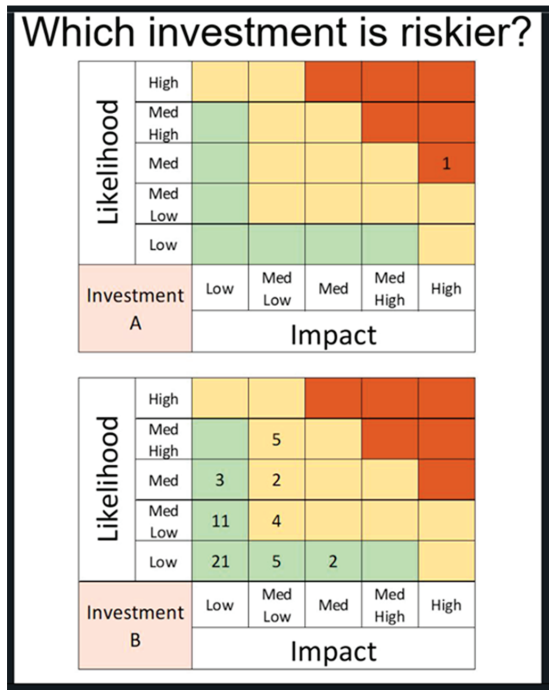


Figure 2. Misuse of models.

There are so many concerns with this question.

- First, a risk is not a single combination of impact and probability. A risk is associated with the range of outcomes of different probabilities; a risk is a curve, not a point. Now, usually when heat-maps plot risks, there is a spot on the map that represents something like a worst credible impact. But that is only for prioritization or to give a vivid summary picture for senior executives or the board of directors. It does not

convey nearly enough information to allow anybody to make any kind of actual decision.

- Second, the two risks in these two maps may not be defined in the same way. It could very well be the case that Investment B involves an array of lower level, more granularly defined risks, and that if you added them all up, they might add up to something that is at least as big as the risk in Investment A. So, that is another set of problems with heat-maps: the definition of the risk, the scope of the risk, and the scale used when evaluating the risks.
- In Investment B, should one of these risk events occur, there may be a domino effect, and once it finishes playing out, there may be a much bigger impact than was identified in the heat-map for Investment A.

Heat-maps have their very limited purposes, but this kind of decision making is definitely not one of them. And yet, dozens of risk ‘experts’ on LinkedIn thought that they could choose between these two alternatives based on these two heat-maps.

Models like heat-maps are only constructed to summarize priorities, encourage thinking, and stimulate conversations. They are not realistic depictions of decision-scenarios. Decision makers have to be reminded that risk models have limitations. Simple models like heat-maps are tremendous simplifications of very complicated realities.

The point is not just about heat-maps. In many cases, there can be real trouble because people trust the models too much. Speculative bubbles are consequences of people trusting valuation models too much. Various banking and financial crises have been about people putting too much faith in metrics, indices numbers, dollars and/or trends (Lowenstein 2000).

3.6. *The Belief That All Risks Are Bad*

Every organization (including hospitals, government agencies, churches, banks, and software companies), regardless of size, scope, or focus, is in the business of putting assets at risk with the expectation of it yielding some kind of return. It is how they make money; it is how they innovate; it is how they increase their influence; it is how they beat their competitors. They all put their assets at risk: their money or their intellectual property or their human capital or their reputation. Otherwise, nothing gets done. Risk is necessary. Risk is good.

However, many people, when they interact with ERM, seem to think that all risk is bad; there is a certain fixation on the downside.

There are two reasons why people perceive that risk is bad. First, it can be what they read, e.g., from regulators, and also from sources such as Sarbanes-Oxley and GRC. These are all about catching and reporting on non-compliance. Much of the upward reporting accomplished seems to focus on the bad things, the exceptions, and the gaps.

Second, it can be about where the ERM group reports organizationally:

- Some organizations combine risk and insurance. Insurance is about the avoidance of loss, i.e., the downside.
- Others have combined ERM with Internal Audit. The role of Internal Audits is basically to identify potential weaknesses in internal controls, i.e., the downside.
- Others place their ERM group in such a manner that they report to a General Counsel. What is the General Counsel’s job? Avoiding legal or commercial risk exposures, i.e., the downside.

Consequently, there is the combined problem of using the wrong tools and reporting to the wrong places. This is why risk appetite is so important. Organizations need to understand their risk appetite and risk tolerance. What risks should be taken? What risks are deliberately chosen? There is no better way to enter into that kind of conversation than getting into an executive discussion about risk appetite. Quail (2021b) presents a practical method for developing a risk appetite that aligns with the strategic ambitions of an organization. One thinks about the different strategic objectives:

- For which ones do we expect that the pathway from where the organization is to where it wants to be is a squiggly/non-linear line, where the organization needs to be responsive and resilient? That suggests a higher risk appetite.
- Which of the strategic objectives are ones for which a small change or volatility in a key performance indicator (KPI) is going to indicate that the organization is lacking in control, and that it would be better to drop everything and figure out what is wrong? That suggests that there is a low risk appetite with respect to that objective.

3.7. A Lack of Role-Clarity

In the early years of ERM, many organizations started their ERM function as a spinoff from Internal Audit. The authors have noticed in recent years that there now seems to be a trend toward recombining these functions. More and more regulators are insisting that the Chief Risk Officer should be an independent assurance function that reports directly to the board of directors. Combining ERM and Internal Audit or assurance, or blurring their distinct roles in this way, can weaken both functions. Internal Audit is no longer independent. Also, ERM is no longer a 'safe' place to discuss risks and the adequacy of mitigation, because the person involved is also an auditor, and part of their job is to report to the board on weaknesses in internal controls and unmitigated risks. So, combining those two functions, or, in fact, combining ERM with any assurance function, is a problem.

There needs to be clarity in the organization as to exactly what ERM is to do. It is not just about producing reports for the board. It is not about providing assurance that the organization is in compliance with SOX. It is not about blowing the whistle when there is a risk about which the organization may think that it might be excessive. It is about helping to stimulate conversations in order to allow management to make better decisions about where the organization allocates resources, what risks to accept, and what risks won't be accepted. It is about building tools and models, all the while clearly communicating their limitations, to help support better decision making. For further reading on ERM and emerging roles in the implementation, refer to Mikes (2021) and Fraser et al. (2021b).

4. Conclusions and Future Directions

In this *JRFM* Communication article, we highlight past research and then present critical examples showing why many organizations fail or flounder in their attempts to implement ERM. Drawing on the extensive experience of an experienced risk manager and consultant, an ex-Chief Risk Officer, and an academic who has researched ERM extensively, we present both the causes and practical solutions in order to assist risk managers in being successful in their implementation of ERM. Numerous surveys show that the successful implementation of ERM trails the expectations of senior management, boards, and regulators (see, for example, Beasley and Branson 2022). This article provides specific practical explanations of the reasons for frequent failures, as well as simple, effective techniques and guidance on how to improve the chances of success in implementing ERM. We provide references for additional guidance on the ERM challenges presented.

ERM is still evolving, and new techniques and research are being explored on a regular basis. Organizations will continue to face business challenges, and risks will continue to evolve as the world becomes more complex. The world is now facing new and fast-evolving risks such as climate risk and the adverse effects of new technologies. There will continue to be unresolved risk issues. ERM will remain key in helping organizations adapt and change in order to remain resilient.

This article should be of interest to organizations implementing ERM, to academics teaching ERM, and to risk professionals desiring to learn more about this evolving process.

Author Contributions: Conceptualization, J.F., R.Q. and B.S.; methodology, J.F. and R.Q.; validation, J.F., R.Q. and B.S.; formal analysis, J.F. and R.Q.; investigation, J.F. and R.Q.; resources, J.F., R.Q. and B.S.; supervision, B.S.; writing—original draft preparation, J.F.; writing—review and editing, J.F., R.Q. and B.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: Author Rob Quail is employed by the company Robert Quail Consulting. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Aabo, Tom, John R. S. Fraser, and Betty J. Simkins. 2005. The rise and transformation of the chief risk officer: A success story on enterprise risk management. *Journal of Applied Corporate Finance* 17: 18–31. [CrossRef]
- Alviniussen, Alf, and Hakan Jankensgård. 2015. Chapter 24 Value and risk: Enterprise risk management at Statoil. In *Implementing Enterprise Risk Management: Case Studies and Best Practices*. Edited by John Fraser, Betty J. Simkins and Kristina Narvaez. Hoboken: John Wiley and Sons, Inc.
- Beasley, Mark, and Bruce Branson. 2022. *The State of Risk Oversight: An Overview of Enterprise Risk Practices*, 13rd ed. Raleigh: North Carolina State University Enterprise Risk Management Initiative.
- COSO (Committee of Sponsoring Organizations of the Treadway Commission). 2017. *Enterprise Risk Management—Integrated Framework*. COSO.
- Crickette, Grace. 2015. Chapter 5 ERM in practice at the University of California Health System. In *Implementing Enterprise Risk Management: Case Studies and Best Practices*. Edited by John Fraser, Betty J. Simkins and Kristina L. Narvaez. Hoboken: John Wiley and Sons, Inc.
- Fraser, John R. S. 2016. The role of the board in risk management oversight. In *Handbook of Corporate Governance*. Edited by Richard Leblanc. Hoboken: John Wiley & Sons.
- Fraser, John R. S., Betty J. Simkins, and Kristina L. Narvaez, eds. 2015. *Implementing Enterprise Risk Management: Case Studies and Best Practices*. Hoboken: John Wiley and Sons, Inc.
- Fraser, John R. S., Rob Quail, and Betty J. Simkins. 2021a. COVID-19: The Risk Management Part Is Unfinished, CFO (January 21). Available online: <https://www.cfo.com/corporate-finance/2021/01/covid-19-the-risk-management-part-is-unfinished-2699/> (accessed on 8 November 2023).
- Fraser, John R. S., Rob Quail, and Betty J. Simkins. 2021b. The history of enterprise risk management at Hydro One Inc. *Journal of Risk and Financial Management* 14: 373. [CrossRef]
- Fraser, John R. S., Rob Quail, and Betty J. Simkins, eds. 2021c. *Enterprise Risk Management: Today's Leading Research and Best Practices for Tomorrow's Executives*, 2nd ed. Hoboken: John Wiley and Sons, Inc.
- International Standards Organization (ISO). 2018. *Risk Management—Guidelines*. ISO 31000. Geneva: International Standards Organization.
- Ismail, Mohamed. 2021. Chapter 24 Organizational Decision Making. In *Enterprise Risk Management: Today's Leading Research and Best Practices for Tomorrow's Executives*, 2nd ed. Edited by John Fraser, Rob Quail and Betty J. Simkins. Hoboken: John Wiley and Sons, Inc., pp. 459–72.
- Kaplan, Robert S., and Anette Mikes. 2012. Managing risks: A new framework. *Harvard Business Review* 90: 48–60.
- Kloman, Felix, and John R. S. Fraser. 2021. Chapter 2 A brief history of risk management. In *Enterprise Risk Management: Today's Leading Research and Best Practices for Tomorrow's Executives*. Edited by John Fraser, Rob Quail and Betty J. Simkins. Hoboken: John Wiley and Sons, Inc., pp. 23–36.
- Lam, James. 2014. *Enterprise Risk Management: From Incentives to Controls*, 2nd ed. Hoboken: John Wiley and Sons, Inc.
- Lam, James. 2017. *Implementing Enterprise Risk Management: From Methods to Applications*. Hoboken: John Wiley and Sons, Inc.
- Lange, David R., and Betty J. Simkins. 2021. Chapter 2 How to teach enterprise risk management: A Learner-centered activities approach. In *Enterprise Risk Management: Today's Leading Research and Best Practices for Tomorrow's Executives*. Edited by John Fraser, Rob Quail and Betty J. Simkins. Hoboken: John Wiley and Sons, Inc.
- Lowenstein, Roger. 2000. *When Genius Failed: The Rise and Fall of Long-Term Capital Management*. New York: Random House.
- Mikes, Anette. 2008. *Enterprise Risk Management at Hydro One (A)*. Harvard Business School Case Study 109-001, July 2008 (Revised January 2012). Brighton: Harvard Business School Publishing. Available online: <https://www.hbs.edu/faculty/Pages/item.aspx?num=36160> (accessed on 1 June 2022).
- Mikes, Anette. 2021. Chapter 8 Becoming the lap bearer: The emerging roles of the chief risk officer. In *Enterprise Risk Management: Today's Leading Research and Best Practices for Tomorrow's Executives*. Edited by John Fraser, Rob Quail and Betty J. Simkins. Hoboken: John Wiley and Sons, Inc.
- Nagumo, Takehiko. 2005. *Aligning Enterprise Risk Management with Strategy through the BSC: The Bank of Tokyo-Mitsubishi Approach, Balanced Scorecard Report*. Reprint No. B0509D, September–October: 1–6. Brighton: Harvard Business School Publishing.
- Narvaez, Kristina L. 2011. *Success Stories: Public Entities Adopt ERM Best Practices*. Alexandria: Public Entity Risk Institute.
- Pagach, Donald, and Heather Pascanik. 2021. Chapter 39 A review of academic research on enterprise risk management. In *Enterprise Risk Management: Today's Leading Research and Best Practices for Tomorrow's Executives*. Edited by John Fraser, Rob Quail and Betty J. Simkins. Hoboken: John Wiley and Sons, Inc.

- Quail, Rob. 2021a. Chapter 19 How to plan and run a risk management workshop. In *Enterprise Risk Management: Today's Leading Research and Best Practices for Tomorrow's Executives*. Edited by John Fraser, Rob Quail and Betty J. Simkins. Hoboken: John Wiley and Sons, Inc.
- Quail, Rob. 2021b. Chapter 23 Risk appetite. In *Enterprise Risk Management: Today's Leading Research and Best Practices for Tomorrow's Executives*, 2nd ed. Edited by John Fraser, Rob Quail and Betty J. Simkins. Hoboken: John Wiley and Sons, Inc., pp. 459–72.
- Quail, Rob. 2021c. Chapter 33. Risk management and outsourcing. In *Enterprise Risk Management: Today's Leading Research and Best Practices for Tomorrow's Executives*. Edited by John Fraser, Rob Quail and Betty J. Simkins. Hoboken: John Wiley and Sons, Inc.
- Sarbanes-Oxley Act. 2002. Senate and House of Representatives of the United States of America in Congress. Available online: [https://www.congress.gov/bill/107th-congress/house-bill/3763#:~:text=Sarbanes-Oxley%20Act%20of%202002%20-%20Title%20I:%20Public%20Company,3\)%20inspect,%20investigate,%20and](https://www.congress.gov/bill/107th-congress/house-bill/3763#:~:text=Sarbanes-Oxley%20Act%20of%202002%20-%20Title%20I:%20Public%20Company,3)%20inspect,%20investigate,%20and) (accessed on 8 November 2023).
- Schwartz, Peter. 1991. *The Art of the Long View: Planning for the Future in an Uncertain World*. New York: Doubleday.
- Stonebrook, Ian. 2021. How Sportswear Sold Streetball. *Boardroom*. August 12. Available online: <https://boardroom.tv/sportswear-streetball-and1/> (accessed on 1 June 2022).
- Taleb, Nassim. 2010. *The Black Swan: Second Edition: The Impact of the Highly Improbable*. New York: Random House.
- Toneguzzo, Joseph P. 2021. How to allocate resources based on risk. In *Enterprise Risk Management: Today's Leading Research and Best Practices for Tomorrow's Executives*. Edited by John Fraser, Rob Quail and Betty J. Simkins. Hoboken: John Wiley and Sons, Inc., chap. 219.
- Warner, Larry. 2015. Chapter 3 ERM at Mars, Incorporated: ERM for Strategy and Operations. In *Implementing Enterprise Risk Management: Case Studies and Best Practices*. Edited by John Fraser, Betty J. Simkins and Kristina L. Navaez. Hoboken: John Wiley and Sons, Inc.
- Wilkinson, Angela, and Roland Kupers. 2013. Living in the futures. *Harvard Business Review* 91: 118–27.
- Winters, Mike. 2021. Chapter 36. Managing risk associated with project delivery: A how to guide. In *Enterprise Risk Management: Today's Leading Research and Best Practices for Tomorrow's Executives*. Edited by John Fraser, Rob Quail and Betty J. Simkins. Hoboken: John Wiley and Sons, Inc.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

Monte Carlo Sensitivities Using the Absolute Measure-Valued Derivative Method

Mark Joshi ^{1,†}, Oh Kang Kwon ^{2,*} and Stephen Satchell ^{2,3}

¹ Centre for Actuarial Studies, University of Melbourne, Parkville, VIC 3010, Australia

² Discipline of Finance, Codrington Building (H69), The University of Sydney, Sydney, NSW 2006, Australia; ses999gb@yahoo.co.uk or stephen.satchell@sydney.edu.au

³ Trinity College, University of Cambridge, Cambridge CB2 1TQ, UK

* Correspondence: ohkang.kwon@sydney.edu.au

† This paper is dedicated to the memory of Mark Joshi who passed away unexpectedly.

Abstract: Measure-valued differentiation (MVD) is a relatively new method for computing Monte Carlo sensitivities, relying on a decomposition of the derivative of transition densities of the underlying process into a linear combination of probability measures. In computing the sensitivities, additional paths are generated for each constituent distribution and the payoffs from these paths are combined to produce sample estimates. The method generally produces sensitivity estimates with lower variance than the finite difference and likelihood ratio methods, and can be applied to discontinuous payoffs in contrast to the pathwise differentiation method. However, these benefits come at the expense of an additional computational burden. In this paper, we propose an alternative approach, called the absolute measure-valued differentiation (AMVD) method, which expresses the derivative of the transition density at each simulation step as a single density rather than a linear combination. It is computationally more efficient than the MVD method and can result in sensitivity estimates with lower variance. Analytic and numerical examples are provided to compare the variance in the sensitivity estimates of the AMVD method against alternative methods.

Keywords: measure-valued differentiation; sensitivities; Greeks; Monte Carlo simulation

JEL Classification: D52; D81; G11

Citation: Joshi, Mark, Oh Kang Kwon, and Stephen Satchell. 2023. Monte Carlo Sensitivities Using the Absolute Measure-Valued Derivative Method. *Journal of Risk and Financial Management* 16: 509. <https://doi.org/10.3390/jrfm16120509>

Academic Editor: Thanasis Stengos

Received: 7 November 2023

Revised: 4 December 2023

Accepted: 5 December 2023

Published: 8 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sensitivities play a key role in the assessment and the management of risks associated with financial derivatives and must be computed along with prices as part of sound risk management practice. Given that it is not possible to compute the sensitivities let alone the prices of many financial derivatives analytically, they are usually computed numerically, with Monte Carlo simulation being the most widely used technique. The associated sensitivities are then computed by finite difference, pathwise differentiation, or likelihood ratio methods, each with its own set of limitations. In particular, the finite difference method is *biased* and sensitivity estimates have relatively large variances, the pathwise derivative method cannot be applied to products with discontinuous payoffs such as digitals, and the likelihood ratio method can be subject to unbounded variance when the associated score function is not absolutely continuous with respect to the sensitivity parameter. For a detailed discussion of these methods and their limitations, refer to Glasserman (2004), chapter 7.

There has been extensive work on regularizing the pathwise method to enable it to work with discontinuities. Such work tends either to involve differentiating the discontinuity directly, which yields a delta distribution, or to utilize smoothing to avoid this differentiation. Joshi and Kaith (2004), Rott and Fries (2005), and Brace (2007) explore the explicit integration of the delta distribution, which requires at least one extra simulation

that must be handcrafted to the payoff. Liu and Hong (2011) and Lyuu and Teng (2011) reduce the degree of handcrafting by employing kernel smoothing and importance sampling, respectively. Chan and Joshi (2015) modify the simulation scheme as a function of the perturbation variable so that the pathwise payoff, as a function of perturbation size, is smooth. It is then possible to apply the pathwise method at the cost of gaining a small likelihood ratio term. All these methods have the drawback that the methodology is adapted to the singularities of the payoff. This renders them much harder to implement generically than the likelihood ratio and pathwise methods.

Despite their importance, there is a relative absence of new techniques for computing Monte Carlo sensitivities, and recent advances have focused instead on improving the computational efficiency of existing techniques. For example, Capriotti and Giles (2018) apply adjoint algorithmic differentiation to speed up pathwise sensitivity calculations, while Xiang and Wang (2022) apply low discrepancy sequences to achieve variance reduction for American option sensitivities.

A notable exception is the measure-valued differentiation (MVD) method considered, for example, in Heidergott et al. (2008) and Pflug and Thoma (2016). The MVD method does not differentiate the payoff and so discontinuities do not cause difficulties, and it has been shown to produce sensitivity estimates with lower variance than those computed using the finite difference and the likelihood ratio (LR) methods. It is also generic in the sense that it is not adapted to the payoff in any way, and so the same paths can be used to compute the sensitivities of many derivatives on the same underlying framework. In the MVD approach, the derivative of the transition density is recognized as a signed measure and decomposed as the difference of two proper measures. The simulation at each step is then branched according to the two constituent measures to produce multiple MVD paths. Sensitivity estimates from these branched paths are then recombined to produce one MVD sensitivity sample. It is shown that this approach generally results in sensitivity estimates with lower variance than alternative methods, albeit at the expense of a greater computational burden due to the branching of the paths. Further variance reduction is possible by coupling, or correlating, each pair of the branched paths.

In this paper, we introduce an alternative to the MVD method, called the *absolute measure-valued differentiation* (AMVD) method, which decomposes the density derivative as the product of its sign and its absolute value. This method shares many of the properties of the MVD method, such as the ability to handle discontinuous payoffs, but instead of initiating two additional paths at each simulation step of the original path, it initiates only one additional subpath. It is found that the AMVD method gives vega estimates with lower variance than the MVD method for vanilla and digital calls in the Black–Scholes model. Moreover, the AMVD vega estimates have lower variance than the pathwise differentiation method in certain situations, such as the case of short-dated deep in-the-money vanilla calls. A consideration of vegas for the more exotic double-barrier options revealed that the relative performances of the LR, MVD, and AMVD methods depend, in general, on the positions of the barriers and the way each method distributes densities relative to the barriers.

Applying the AMVD method requires the generation of nonstandard random variates,¹ and for the examples considered in this paper, we provide a highly efficient implementation of the inverse transform method for this purpose using Newton's method with a careful selection of the initial point. The time taken to generate each vega subpath for the AMVD method is approximately equal to the time taken to generate the corresponding MVD method using the efficient generation of the double-sided Boltzmann–Maxwell distribution introduced in Heidergott et al. (2008).

The remainder of the paper is structured as follows: the measure-valued differentiation method is briefly reviewed in Section 2, and the absolute measure-valued differentiation method is presented in Section 3. This is followed by Section 4 in which the notation and key identities for the remainder of the paper are presented. Sensitivities for vanilla calls and digital calls are computed analytically in Sections 5 and 6, respectively. A more complex

example of the double knock-out barrier option is investigated numerically in Section 7, and the paper concludes with Section 8.

2. Brief Review of Measure-Valued Differentiation

In this section, we provide a heuristic overview of the MVD method and refer the interested reader to Heidergott et al. (2008) and Pflug and Thoma (2016) for the details.

Let X be an $\mathbb{R}$ -valued random variable, and let $\phi(x; \theta)$ be the probability density function of X with θ as one of its parameters. Then, given a payoff, $f: \mathbb{R} \rightarrow \mathbb{R}$, associated with a financial derivative on X , the likelihood ratio (LR) sensitivity with respect to the parameter θ is given by

$$\Delta_{\theta}(f) = \int_{\mathbb{R}} f(x) \frac{\partial \phi(x; \theta)}{\partial \theta} dx \quad (1)$$

ignoring discounting. At this point, the likelihood ratio method rewrites the above expression as

$$\Delta_{\theta}^{\text{LR}}(f) = \int_{\mathbb{R}} \frac{\partial \ln \phi(x; \theta)}{\partial \theta} f(x) \cdot \phi(x; \theta) dx, \quad (2)$$

and computes the sensitivity as the expectation of the *modified* payoff obtained by scaling $f(x)$ by the *score function* $\partial \ln \phi(x; \theta) / \partial \theta$.

Now, suppose $\varphi(x) = \partial \phi(x; \theta) / \partial \theta$ is a signed measure on $\mathbb{R}$, and that it decomposes as the difference

$$\varphi(x) = \varphi_1(x) - \varphi_2(x), \quad (3)$$

where φ_1 and φ_2 are proper measures on $\mathbb{R}$. Then, under the MVD method, the sensitivity with respect to θ is computed as the difference

$$\Delta_{\theta}^{\text{MVD}}(f) = \mu(\varphi_1) \int_{\mathbb{R}} f(x) \frac{\varphi_1(x)}{\mu(\varphi_1)} dx - \mu(\varphi_2) \int_{\mathbb{R}} f(x) \frac{\varphi_2(x)}{\mu(\varphi_2)} dx, \quad (4)$$

where $\mu(\varphi_i) = \int_{\mathbb{R}} \varphi_i(x) dx > 0$. The two summands are computed separately, although they may be coupled to reduce the total variance. It should be noted that $\bar{\varphi}_i(x) = \varphi_i(x) / \mu(\varphi_i)$, for $1 \leq i \leq 2$, are probability measures on $\mathbb{R}$, and so the decomposition in (4) provides the basis for a numerical approximation of the sensitivity using, for example, a Monte Carlo simulation in which the integrals in (4) are computed numerically using random draws from the distributions determined by $\bar{\varphi}_i(x)$.

The above description of the MVD method applies to path-independent European derivatives. For the more general case, suppose that a Monte Carlo simulation, with simulation times $0 = t_0 < t_1 < \dots < t_n = T$, is used to value a given derivative with expiry T . If the underlying process is assumed to be Markovian, then the probability density of a simulated path is given by the product

$$\phi(x_0, x_1, \dots, x_n; \theta) = \prod_{i=1}^n \phi_i(x_i | x_{i-1}; \theta), \quad (5)$$

where $\phi_i(x_i | x_{i-1}; \theta)$ is the transition density from the t_{i-1} to t_i conditional starting from x_{i-1} at t_{i-1} . For notational convenience, write $\phi(x; \theta) = \phi(x_0, x_1, \dots, x_n; \theta)$, $\phi_i(x_i; \theta) = \phi_i(x_i | x_{i-1}; \theta)$, and define

$$\partial_{i,\theta} \phi(x; \theta) = \phi_1(x_1; \theta) \cdots \phi_{i-1}(x_{i-1}; \theta) \frac{\partial \phi_i(x_i; \theta)}{\partial \theta} \phi_{i+1}(x_{i+1}; \theta) \cdots \phi_n(x_n; \theta) \quad (6)$$

The likelihood ratio sensitivity of a general payoff, $f(\mathbf{x}) = f(x_0, x_1, x_2, \dots, x_n)$, with respect to the parameter, θ , is then given by

$$\Delta_{\theta}^{\text{LR}}(f) = \sum_{i=1}^n \int_{\mathbb{R}^n} f(\mathbf{x}) \partial_{i,\theta} \phi(\mathbf{x}; \theta) d\mathbf{x}. \quad (7)$$

Once again, if it is assumed that the derivatives

$$\varphi_i(x_i) = \frac{\partial \phi_i(x_i, \theta)}{\partial \theta} \quad (8)$$

are signed measures on $\mathbb{R}$, and decompose as differences

$$\varphi_i(x_i) = \varphi_{i,1}(x_i) - \varphi_{i,2}(x_i), \quad (9)$$

where $\varphi_{i,j}(x)$ are proper measures, then the MVD estimate of the sensitivity is given by

$$\Delta_{\theta}^{\text{MVD}}(f) = \sum_{i=1}^n \int_{\mathbb{R}^n} f(\mathbf{x}) \partial_{i,\theta}^{\text{MVD}} \phi(\mathbf{x}; \theta) d\mathbf{x}, \quad (10)$$

where $\partial_{i,\theta}^{\text{MVD}} \phi(\mathbf{x}; \theta)$ is obtained from $\partial_{i,\theta} \phi(\mathbf{x}; \theta)$ by substituting the derivative term, $\partial \phi_i(x_i; \theta) / \partial \theta$, in (6) with its decomposition

$$\frac{\partial \phi_i(x_i; \theta)}{\partial \theta} = \mu(\varphi_{i,1}) \frac{\varphi_{i,1}(x_i)}{\mu(\varphi_{i,1})} - \mu(\varphi_{i,2}) \frac{\varphi_{i,2}(x_i)}{\mu(\varphi_{i,2})}, \quad (11)$$

where $\mu(\varphi_{i,j}) = \int_{\mathbb{R}} \varphi_{i,j}(x_i) dx_i$.

The implementation of the MVD method begins with the simulation, in the usual manner, of the underlying asset price path. Then, for each $1 \leq i \leq n$, the i -th step is branched into two by drawing from distributions corresponding to the probability densities, $\bar{\varphi}_{i,j}(x) = \varphi_{i,j}(x) / \mu(\varphi_{i,j})$, for $1 \leq j \leq 2$. The remaining steps in the two branched paths are then resimulated by drawing from the original distributions, $\phi_j(x_j; \theta)$ for $i+1 \leq j \leq n$. Thus, the *full* MVD method generates a set of $2n$ paths for each regular path and is consequently more computationally intensive than, for example, the likelihood ratio method.² In Pflug and Thoma (2016), less computationally demanding approximations to the full MVD method are proposed in which branching is performed only in a small subset of the simulation steps.

3. Absolute Measure-Valued Differentiation

An important factor taken into consideration when deciding on the form of the decomposition of the transition density in (9) is the additional computation burden associated with the need to generate random variates from the component distributions. The decomposition chosen in Heidergott et al. (2008) and Pflug and Thoma (2016) is due, at least in part, to ensuring that one of the components coincides with the distribution *already* used in generating the underlying process. The remaining factor, however, will generally be from a different distribution.

The main motivation behind the *absolute measure-valued differentiation* (AMVD) method is improve the computational efficiency of the MVD method by requiring the introduction of only one additional distribution and branching off only one path at each simulation step, while retaining the key features of the MVD method, viz., not requiring the payoff derivative, and producing sensitivity estimates with low variances. Using the notation from the previous section, the AMVD estimate for the sensitivity of a payoff f with respect to the parameter θ is defined by

$$\Delta_{\theta}^{\text{AMVD}}(f) = \sum_{i=1}^n \int_{\mathbb{R}^n} f(x) \varepsilon(\varphi_i(x_i)) \mu(|\varphi_i|) \frac{\partial_{|\varphi_i|, \theta}^{\text{MVD}} \phi(x; \theta)}{\mu(|\varphi_i|)} dx, \quad (12)$$

where $\varepsilon(\varphi_i(x_i)) = \mathbf{1}_{\{\varphi_i(x_i) \geq 0\}} - \mathbf{1}_{\{\varphi_i(x_i) < 0\}}$, $\mu(|\varphi_i|) = \int_{\mathbb{R}} |\varphi_i(x_i)| dx_i$, and

$$\partial_{|\varphi_i|, \theta}^{\text{MVD}} \phi(x; \theta) = \phi_1(x_1; \theta) \cdots \phi_{i-1}(x_{i-1}; \theta) |\varphi_i(x_i)| \phi_{i+1}(x_{i+1}; \theta) \cdots \phi_n(x_n; \theta). \quad (13)$$

It follows from $\varepsilon(\varphi_i(x_i)) |\varphi_i(x_i)| = \varphi_i(x_i)$ that $\Delta_{\theta}^{\text{AMVD}}(f)$ is an unbiased estimator for $\Delta_{\theta}(f)$. Moreover, although the AMVD method introduces what appears to be a problematic term, viz., $\varepsilon(\varphi_i(x_i))$, since $\varepsilon(\varphi_i(x_i)) = \pm 1$, and these terms are squared when computing the associated variance, they do not adversely affect the AMVD variance.

4. Notation, Model, and Useful Identities

In situations where it applies, the pathwise differentiation (PW) method usually gives sensitivity estimates with lower variances than other methods, and it was shown in Pflug and Thoma (2016) that the MVD sensitivities have lower variance than the corresponding finite difference estimates. In subsequent sections, we will compare the variances of sensitivity estimates computed using the LR, MVD, AMVD, and PW methods. Although the analysis could be performed for any sensitivity and with any underlying model, we will focus on delta and vega under the Black and Scholes (1973) model. This allows us to obtain, analytically, the variances of deltas and vegas for vanillas and digitals, and avoid the consideration of more complex models that do not contribute, in any substantial way, to the analysis of the relative performance of the various approaches for computing sensitivities.

4.1. Black–Scholes Delta and Vega

Let r be the risk-free rate, S_t the price of the underlying asset at time t , and σ the asset volatility. Moreover, let $0 = t_0 < t_1 < \cdots < T_n = T$ be the sequence of Monte Carlo simulation times, and, for notational convenience, define

$$S_i = S_{t_i}, \quad x_i = \ln S_i, \quad \text{and} \quad \Delta_j = t_{j+1} - t_j, \quad (14)$$

where $0 \leq i \leq n$ and $0 \leq j \leq n-1$. The Euler–Maruyama discretization scheme for the Black–Scholes model

$$x_{i+1} = x_i + \left(r - \frac{1}{2}\sigma^2\right)\Delta_i + \sigma\sqrt{\Delta_i}\xi_i \quad (15)$$

is unbiased, where $\xi_i \sim \mathcal{N}(0, 1)$ is a normally distributed random variable. If we define

$$\mu_i(\sigma) = \left(r - \frac{1}{2}\sigma^2\right)\Delta_{i-1} \quad \text{and} \quad \varsigma_i(\sigma) = \sigma\sqrt{\Delta_{i-1}}, \quad (16)$$

for $1 \leq i \leq n$, then the transition density, $\phi_i(x_i; \sigma)$, is given by

$$\phi_i(x_i; \sigma) = \frac{1}{\varsigma_i(\sigma)\sqrt{2\pi}} \exp\left[-\frac{(x_i - x_{i-1} - \mu_i(\sigma))^2}{2\varsigma_i^2(\sigma)}\right], \quad (17)$$

so that $x_i \sim \mathcal{N}(x_{i-1} + \mu_i(\sigma), \varsigma_i^2(\sigma))$. Straight-forward calculations give

$$\phi_1(x_i; S_0) = \frac{\partial \phi_1(x_1; S_0)}{\partial S_0} = \frac{1}{S_0 \varsigma_1(\sigma)} \xi_1 \phi(\xi_1), \quad (18)$$

$$\phi_i(x_i; S_0) = 0, \quad 2 \leq i \leq n, \quad (19)$$

$$\phi_i(x_i; \sigma) = \frac{\partial \phi_i(x_i; \sigma)}{\partial \sigma} = \frac{1}{\sigma} (\xi_i^2 - \varsigma_i \xi_i - 1) \phi(\xi_i), \quad 1 \leq i \leq n, \quad (20)$$

where ξ_i is the standard normal random variate in (15) and $\phi(\cdot)$ is the standard normal density. The decompositions chosen for the MVD method in Pflug and Thoma (2016) are

$$\varphi_1^{\text{MVD}}(x_i; S_0) = \frac{1}{S_0 \xi_1(\sigma)} \xi_1 \phi(\xi_1) \mathbf{1}_{\{\xi_1 \geq 0\}} - \frac{1}{S_0 \xi_1(\sigma)} (-\xi_1) \phi(\xi_1) \mathbf{1}_{\{\xi_1 \leq 0\}}, \quad (21)$$

$$\varphi_i^{\text{MVD}}(x_i; \sigma) = \frac{1}{\sigma} \xi_i^2 \phi(\xi_i) - \frac{\xi_i(\sigma)}{\sigma} \xi_i \phi(\xi_i) \mathbf{1}_{\{\xi_i \geq 0\}} + \frac{\xi_i(\sigma)}{\sigma} (-\xi_i) \phi(\xi_i) \mathbf{1}_{\{\xi_i \leq 0\}} - \frac{1}{\sigma} \phi(\xi_i), \quad (22)$$

where $2 \leq i \leq n$. Both terms in (21) correspond to the density of the Rayleigh distribution, while in (22), the first term corresponds to the double-sided Maxwell–Boltzmann distribution, the second and third terms correspond to the Rayleigh distribution, and the last term is the standard normal density. Note that the linear terms in (18) and (20) were decomposed into two Rayleigh density terms since they define *signed* measures rather than proper measures on $\mathbb{R}$. The decompositions that will be used in the AMVD method are

$$\varphi_1^{\text{AMVD}}(x_i; S_0) = \frac{1}{S_0 \xi_1(\sigma)} |\xi_1 \phi(\xi_1)| \phi(\xi_1), \quad (23)$$

$$\varphi_i^{\text{AMVD}}(x_i; \sigma) = \frac{1}{\sigma} \varepsilon \left(\xi_i^2 - \xi_i(\sigma) \xi_i - 1 \right) \left| \xi_i^2 - \xi_i(\sigma) \xi_i - 1 \right| \phi(\xi_i). \quad (24)$$

4.2. Notation and Identities

For any $n \in \mathbb{N}_+$, define polynomials, $p_n(x)$, by $p_0(x) = 1$ and for $n \geq 1$,

$$p_n(x) = x^n - (n-1)x^{n-2}. \quad (25)$$

Moreover, given any polynomial, $p(x)$, of degree n , and $a \in \mathbb{R}$, let $\alpha_i(p, a) \in \mathbb{R}$ be the coefficients in the decomposition

$$p(x+a) = \sum_{i=0}^n \alpha_i(p, a) p_i(x), \quad (26)$$

which are welldefined and unique since $\{p_i(x) \mid 0 \leq i \leq n\}$ is a basis for the space of polynomials of degree at the highest n . For the purposes of computing the sensitivity variances, we note that

$$(x+a)^2 = 1 \cdot (x^2 - 1) + 2a \cdot x + (a^2 + 1) \cdot 1, \quad (27)$$

so that

$$\alpha_0(x^2, a) = a^2 + 1, \quad \alpha_1(x^2, a) = 2a, \quad \text{and} \quad \alpha_2(x^2, a) = 1. \quad (28)$$

If for any given $\zeta \in \mathbb{R}$, we define the quadratic $q_\zeta(x)$, by

$$q_\zeta(x) = x^2 - \zeta x - 1, \quad (29)$$

then

$$q_\zeta(x+a) = 1 \cdot (x^2 - 1) + (2a - \zeta) \cdot x + a(a - \zeta) \cdot 1, \quad (30)$$

and so

$$\alpha_0(q_\zeta(x), a) = a(a - \zeta), \quad \alpha_1(q_\zeta(x), a) = 2a - \zeta, \quad \text{and} \quad \alpha_2(q_\zeta(x), a) = 1. \quad (31)$$

Next, squaring $q_\zeta(x)$ gives

$$\begin{aligned} q_\zeta^2(x+a) &= 1 \cdot (x^4 - 3x^2) + 2(2a - \zeta) \cdot (x^3 - 2x) \\ &\quad + (6a^2 - 6\zeta a + \zeta^2 + 1) \cdot (x^2 - 1) \\ &\quad + (4a^3 - 6a^2\zeta + 2a(\zeta^2 + 2) - 2\zeta) \cdot x \\ &\quad + (a^4 - 2\zeta a^3 + (\zeta^2 + 4)a^2 - 4\zeta a + \zeta^2 + 2) \cdot 1, \end{aligned} \quad (32)$$

and so $\alpha_4(q_\zeta^2(x), a) = 1$, $\alpha_3(q_\zeta^2(x), a) = 2(2a - \zeta)$, and

$$\alpha_0(q_\zeta^2(x), a) = a^4 - 2\zeta a^3 + (\zeta^2 + 4)a^2 - 4\zeta a + \zeta^2 + 2, \quad (33)$$

$$\alpha_1(q_\zeta^2(x), a) = 4a^3 - 6a^2\zeta + 2a(\zeta^2 + 2) - 2\zeta, \quad (34)$$

$$\alpha_2(q_\zeta^2(x), a) = 6a^2 - 6\zeta a + \zeta^2 + 1. \quad (35)$$

Finally, for any $i \in \mathbb{N}$, define

$$\chi_i = (i-1) \vee 1 = \max(i-1, 1), \quad (36)$$

and given $a \in \mathbb{R}$, $l \leq u \in \mathbb{R}$, and a polynomial $p(x)$ of degree n , define

$$\begin{aligned} \mathcal{I}(p, a; l, u) &= \alpha_0(p, a)[\Phi(u-a) - \Phi(l-a)] \\ &\quad + \phi(l-a) \sum_{i=1}^n \chi_i \alpha_i(p, a)(l-a)^{i-1} \\ &\quad - \phi(u-a) \sum_{i=1}^n \chi_i \alpha_i(p, a)(u-a)^{i-1}. \end{aligned} \quad (37)$$

In (37), we agree to set $\Phi(-\infty) = \phi(\pm\infty) = 0$ and $\Phi(\infty) = 1$. Note that the definition of $\mathcal{I}(p, a; l, u)$ was motivated by the identity

$$\int_l^u e^{ax} p(x) \phi(x) dx = e^{\frac{1}{2}a^2} \mathcal{I}(p, a; l, u). \quad (38)$$

It will be seen that the variances of deltas and vegas for vanilla and digital options can be obtained analytically and expressed succinctly in terms of $\mathcal{I}(p, a; l, u)$.

5. Vanilla Calls

For vanilla options, it is possible to derive analytic expressions for the variances of deltas and vegas for the LR, MVD, AMVD, and PW methods. We only provide the details for European calls, since the corresponding calculations for puts are entirely similar. Throughout this section, let K be the strike for a European call with expiry T , and let $\zeta = \sigma\sqrt{T}$.

5.1. Vanilla Call Delta

Recall from (18) that if $\Delta_1 = T$, then the derivative of the transition density from $t = 0$ to $t = T$ with respect to S_0 is given by

$$\varphi(x_T; S_0) = \frac{\partial \phi(x_T; S_0)}{\partial S_0} = \frac{1}{S_0 \zeta} \xi \phi(\xi), \quad (39)$$

where $\xi \sim \mathcal{N}(0, 1)$, and for notational clarity, we have written $x_T = x_1$, $\xi = \xi_1$ and $\varphi(x_T; S_0) = \varphi_1(x_1; S_0)$. The likelihood ratio estimator for the call delta is unbiased and computed using the expression

$$\mathbb{E}[\Delta_{S_0}^{\text{LR}}(c)] = \int_{-\xi_-}^{\infty} \frac{e^{-rT}\xi}{S_0\zeta} \left(S_0 e^{(r-\frac{1}{2}\sigma^2)T+\zeta\xi} - K \right) \cdot \phi(\xi) d\xi = \Phi(\xi_+),$$

where

$$\xi_{\pm} = \frac{\ln(S_0/K) + \left(r \pm \frac{1}{2}\sigma^2\right)T}{\zeta}. \quad (40)$$

The variance of the likelihood ratio call delta is given by

$$\begin{aligned} \mathbb{V}[\Delta_{S_0}^{\text{LR}}(c)] &= \frac{e^{\zeta^2}}{\zeta^2} \mathcal{I}(x^2, 2\zeta; -\xi_-, \infty) - \frac{2Ke^{-rT}}{S_0\zeta^2} \mathcal{I}(x^2, \zeta; -\xi_-, \infty) \\ &\quad + \frac{K^2e^{-2rT}}{S_0^2\zeta^2} \mathcal{I}(x^2, 0; -\xi_-, \infty) - \Phi^2(\xi_+). \end{aligned} \quad (41)$$

The expected value of the MVD delta is *unbiased* and is computed using the expression

$$\begin{aligned} \mathbb{E}[\Delta_{S_0}^{\text{MVD}}(c)] &= \int_{-\xi_-}^{\infty} \frac{e^{-rT}}{S_0\zeta\sqrt{2\pi}} \left[S_0 e^{(r-\frac{1}{2}\sigma^2)T+\zeta\xi} - K \right] \cdot \mathbf{1}_{\{\xi \geq 0\}} \sqrt{2\pi} \xi \phi(\xi) d\xi \\ &\quad - \int_{-\xi_-}^{\infty} \frac{e^{-rT}}{S_0\zeta\sqrt{2\pi}} \left[S_0 e^{(r-\frac{1}{2}\sigma^2)T+\zeta\xi} - K \right] \cdot \mathbf{1}_{\{\xi < 0\}} \sqrt{2\pi} (-\xi) \phi(\xi) d\xi, \end{aligned}$$

where the terms are computed *independently*. The coefficient, $\sqrt{2\pi}$, preceding the terms $\pm \xi \phi(\xi)$, is introduced to ensure that the product is a probability density function. The variance of the MVD delta is given by

$$\mathbb{V}[\Delta_{S_0}^{\text{MVD}}(c)] = m_{S_0}^{(2)}(c) - \Phi^2(\xi_-), \quad (42)$$

where

$$\begin{aligned} m_{S_0}^{(2)}(c) &= \frac{e^{\zeta^2}}{\sqrt{2\pi}\zeta^2} \mathcal{I}(x, 2\zeta; 0 \vee -\xi_-, \infty) - \frac{2Ke^{-rT}}{\sqrt{2\pi}S_0\zeta^2} \mathcal{I}(x, \zeta; 0 \vee -\xi_-, \infty) \\ &\quad + \frac{K^2e^{-2rT}}{\sqrt{2\pi}S_0^2\zeta^2} \mathcal{I}(x, 0; 0 \vee -\xi_-, \infty) - \frac{e^{\zeta^2}}{\sqrt{2\pi}\zeta^2} \mathcal{I}(x, 2\zeta; 0 \wedge -\xi_-, 0) \\ &\quad + \frac{2Ke^{-rT}}{\sqrt{2\pi}S_0\zeta^2} \mathcal{I}(x, \zeta; 0 \wedge -\xi_-, 0) - \frac{K^2e^{-2rT}}{\sqrt{2\pi}S_0^2\zeta^2} \mathcal{I}(x, 0; 0 \wedge -\xi_-, 0). \end{aligned} \quad (43)$$

The expected value of the AMVD delta is *unbiased* and is computed using the expression

$$\mathbb{E}[\Delta_{S_0}^{\text{AMVD}}(c)] = \int_{-\xi_-}^{\infty} \varepsilon(\xi) \frac{e^{-rT}}{S_0\zeta} \sqrt{\frac{2}{\pi}} \left[S_0 e^{(r-\frac{1}{2}\sigma^2)T+\zeta\xi} - K \right] \cdot \sqrt{\frac{\pi}{2}} |\xi| \phi(\xi) d\xi,$$

and its variance is given by

$$\mathbb{V}[\Delta_{S_0}^{\text{AMVD}}(c)] = 2m_{S_0}^{(2)}(c) - \Phi^2(\xi_+). \quad (44)$$

It should be noted that although $\mathbb{V}[\Delta_{S_0}^{\text{MVD}}(c)]$ is approximately half the value of $\mathbb{V}[\Delta_{S_0}^{\text{AMVD}}(c)]$, implementing MVD using a Monte Carlo simulation would require two paths for each sample, while the AMVD method requires only one. Finally, the expected value for the PW delta is given by

$$\mathbb{E}[\Delta_{S_0}^{\text{PW}}(c)] = \int_{-\xi_-}^{\infty} e^{-\frac{1}{2}\zeta^2+\zeta\xi} \cdot \phi(\xi) d\xi,$$

with the associated variance

$$\mathbb{V}[\Delta_{S_0}^{\text{PW}}(c)] = e^{\zeta^2} \mathcal{I}(1, 2\zeta; -\zeta_-, \infty) - \Phi^2(\zeta_+) = e^{\zeta^2} \Phi(\zeta_+ + \zeta) - \Phi^2(\zeta_+). \quad (45)$$

The ratios of variances for the vanilla call delta are shown in Figures 1 and 2. It can be seen that the AMVD method gives lower variances than the LR method, but higher variances than the MVD method. Somewhat unexpectedly, the AMVD method results in lower variances than the PW method for short-dated deep in-the-money calls.

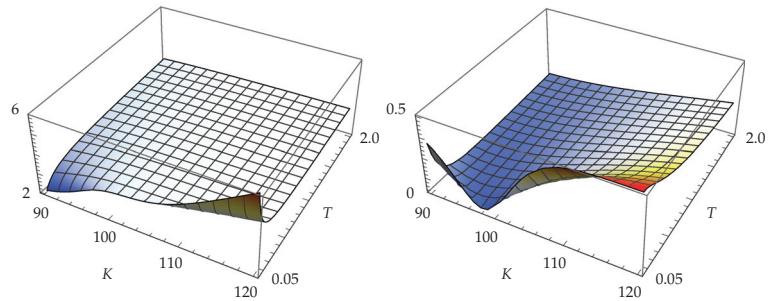


Figure 1. On the left, there is the ratio $\mathbb{V}[\Delta_{S_0}^{\text{LR}}(c)] / \mathbb{V}[\Delta_{S_0}^{\text{AMVD}}(c)]$ and, on the right, there is the ratio $\mathbb{V}[\Delta_{S_0}^{\text{MVD}}(c)] / \mathbb{V}[\Delta_{S_0}^{\text{AMVD}}(c)]$.

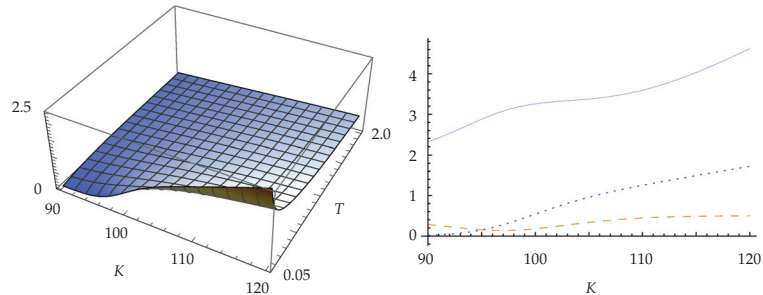


Figure 2. On the left, there is the ratio $\mathbb{V}[\Delta_{S_0}^{\text{PW}}(c)] / \mathbb{V}[\Delta_{S_0}^{\text{AMVD}}(c)]$. On the right, there are variance ratios at $T = 0.1$. The solid curve is $\mathbb{V}[\Delta_{S_0}^{\text{LR}}(c)] / \mathbb{V}[\Delta_{S_0}^{\text{AMVD}}(c)]$, the dashed curve is $\mathbb{V}[\Delta_{S_0}^{\text{MVD}}(c)] / \mathbb{V}[\Delta_{S_0}^{\text{AMVD}}(c)]$, and the dotted curve is $\mathbb{V}[\Delta_{S_0}^{\text{PW}}(c)] / \mathbb{V}[\Delta_{S_0}^{\text{AMVD}}(c)]$.

5.2. Vanilla Call Vega

The derivative of the transition density, $\phi(x_T; \sigma)$, with respect to σ is given by

$$\varphi(x_T; \sigma) = \frac{\partial \phi(x_T; \sigma)}{\partial \sigma} = \frac{1}{\sigma} (\zeta^2 - \zeta \zeta_- - 1) \phi(\zeta), \quad (46)$$

where $\zeta = \sigma \sqrt{T}$ is as defined previously. The likelihood ratio vega for the vanilla call is unbiased and is given by

$$\mathbb{E}[\Delta_{\sigma}^{\text{LR}}(c)] = \int_{-\zeta_-}^{\infty} \frac{e^{-rT}}{\sigma} \left[S_0 e^{(r - \frac{1}{2}\sigma^2)T + \sigma \sqrt{T} \zeta} - K \right] (\zeta^2 - \zeta \zeta_- - 1) \cdot \phi(\zeta) d\zeta,$$

so that $\mathbb{E}[\Delta_\sigma^{\text{LR}}] = S_0\phi(\xi_+)\sqrt{T}$. The associated variance is

$$\begin{aligned} \mathbb{V}[\Delta_\sigma^{\text{LR}}(c)] &= \frac{e^{\xi_-^2} S_0^2}{\sigma^2} \mathcal{I}(q_\xi^2(x), 2\xi; -\xi_-, \infty) - \frac{2e^{-rT} K S_0}{\sigma^2} \mathcal{I}(q_\xi^2(x), \xi; -\xi_-, \infty) \\ &\quad + \frac{e^{-2rT} K^2}{\sigma^2} \mathcal{I}(q_\xi^2(x), 0; -\xi_-, \infty) - S_0^2 \phi^2(\xi_+) T. \end{aligned} \quad (47)$$

The MVD vanilla call vega is unbiased and is computed using the decomposition

$$\begin{aligned} \mathbb{E}[\Delta_\sigma^{\text{MVD}}(c)] &= \int_{-\xi_-}^{\infty} \frac{e^{-rT}}{\sigma} [S_0 e^{(r-\frac{1}{2}\sigma^2)T + \sigma\sqrt{T}\xi} - K] \cdot \xi^2 \phi(\xi) d\xi \\ &\quad - \int_{0 \vee -\xi_-}^{\infty} \frac{e^{-rT} \xi}{\sigma \sqrt{2\pi}} [S_0 e^{(r-\frac{1}{2}\sigma^2)T + \sigma\sqrt{T}\xi} - K] \cdot \sqrt{2\pi} \xi \phi(\xi) d\xi \\ &\quad + \int_{0 \wedge -\xi_-}^0 \frac{e^{-rT} \xi}{\sigma \sqrt{2\pi}} [S_0 e^{(r-\frac{1}{2}\sigma^2)T + \sigma\sqrt{T}\xi} - K] \cdot \sqrt{2\pi} (-\xi) \phi(\xi) d\xi \\ &\quad - \int_{-\xi_-}^{\infty} \frac{e^{-rT}}{\sigma} [S_0 e^{(r-\frac{1}{2}\sigma^2)T + \sigma\sqrt{T}\xi} - K] \cdot \phi(\xi) d\xi, \end{aligned}$$

where it is assumed that the components are uncoupled. It should be noted that the first and last terms are coupled in Heidergott et al. (2008) and Pflug and Thoma (2016) to reduce the resulting variance. The variance associated with the uncoupled MVD vanilla call vega is

$$\begin{aligned} \mathbb{V}[\Delta_\sigma^{\text{MVD}}(c)] &= \frac{e^{\xi_-^2} S_0^2}{\sigma^2} [\mathcal{I}(x^2, 2\xi; -\xi_-, \infty) + \mathcal{I}(1, 2\xi; -\xi_-, \infty)] \\ &\quad - \frac{2e^{-rT} K S_0}{\sigma^2} [\mathcal{I}(x^2, \xi; -\xi_-, \infty) + \mathcal{I}(1, \xi; -\xi_-, \infty)] \\ &\quad + \frac{e^{-2rT} K^2}{\sigma^2} [\mathcal{I}(x^2, 0; -\xi_-, \infty) + \mathcal{I}(1, 0; -\xi_-, \infty)] \\ &\quad + \frac{e^{\xi_-^2} S_0^2 T}{\sqrt{2\pi}} [\mathcal{I}(x, 2\xi; 0 \vee -\xi_-, \infty) - \mathcal{I}(x, 2\xi; 0 \wedge -\xi_-, 0)] \\ &\quad - \frac{2e^{-rT} K S_0 T}{\sqrt{2\pi}} [\mathcal{I}(x, \xi; 0 \vee -\xi_-, \infty) - \mathcal{I}(x, \xi; 0 \wedge -\xi_-, 0)] \\ &\quad + \frac{e^{-2rT} K^2 T}{\sqrt{2\pi}} [\mathcal{I}(x, 0; 0 \vee -\xi_-, \infty) - \mathcal{I}(x, 0; 0 \wedge -\xi_-, 0)] \\ &\quad - S_0^2 \phi^2(\xi_+) T. \end{aligned} \quad (48)$$

The AMVD vanilla call vega is computed using the expression

$$\mathbb{E}[\Delta_\sigma^{\text{AMVD}}(c)] = \int_{-\xi_-}^{\infty} \varepsilon(q_\xi(x)) \frac{e^{-rT} \eta_\xi}{\sigma} [S_0 e^{(r-\frac{1}{2}\sigma^2)T + \xi\xi} - K] \cdot \frac{1}{\eta_\xi} |q_\xi(\xi)| \phi(\xi) d\xi,$$

where

$$\eta_\xi = \mathcal{I}(q_\xi, 0; -\infty, \xi_-) - \mathcal{I}(q_\xi, 0; \xi_-, \xi_+) + \mathcal{I}(q_\xi, 0; \xi_+, \infty), \quad (49)$$

$$\xi_\pm = \frac{1}{2}\xi \pm \sqrt{1 + \frac{1}{4}\xi^2}, \quad (50)$$

and the associated variance is

$$\begin{aligned}
 \mathbb{V}[\Delta_{\sigma}^{\text{AMVD}}(c)] = & \frac{e^{\zeta^2} S_0^2 \eta_{\zeta}}{\sigma^2} [\mathcal{I}(q(x), 2\zeta; \zeta_+ \vee -\zeta_-, \infty) \\
 & - \mathcal{I}(q(x), 2\zeta; (\zeta_- \vee -\zeta_-) \wedge \zeta_+, \zeta_+) \\
 & + \mathcal{I}(q(x), 2\zeta; \zeta_- \wedge -\zeta_-, \zeta_-)] \\
 & - \frac{2e^{-rT} K S_0 \eta_{\zeta}}{\sigma^2} [\mathcal{I}(q(x), \zeta; \zeta_+ \vee -\zeta_-, \infty) \\
 & - \mathcal{I}(q(x), \zeta; (\zeta_- \vee -\zeta_-) \wedge \zeta_+, \zeta_+) \\
 & + \mathcal{I}(q(x), \zeta; \zeta_- \wedge -\zeta_-, \zeta_-)] \\
 & + \frac{e^{-2rT} K^2 \eta_{\zeta}}{\sigma^2} [\mathcal{I}(q(x), 0; \zeta_+ \vee -\zeta_-, \infty) \\
 & - \mathcal{I}(q(x), 0; (\zeta_- \vee -\zeta_-) \wedge \zeta_+, \zeta_+) \\
 & + \mathcal{I}(q(x), 0; \zeta_- \wedge -\zeta_-, \zeta_-)] \\
 & - S_0^2 \phi^2(\zeta_+) T.
 \end{aligned} \tag{51}$$

Finally, the pathwise vega for the vanilla call is given by

$$\mathbb{E}[\Delta_{\sigma}^{\text{PW}}(c)] = \int_{-\zeta_-}^{\infty} (-\sigma T + \sqrt{T} \xi) S_0 e^{-\frac{1}{2}\zeta^2 + \zeta \xi} \phi(\xi) d\xi,$$

with the associated variance

$$\begin{aligned}
 \mathbb{V}[\Delta_{\sigma}^{\text{PW}}(c)] = & e^{\zeta^2} S_0^2 T [\mathcal{I}(x^2, 2\zeta; -\zeta_-, \infty) - 2\zeta \mathcal{I}(x, 2\zeta; -\zeta_-, \infty) \\
 & + \zeta^2 \mathcal{I}(1, 2\zeta; -\zeta_-, \infty)] - S_0^2 \phi^2(\zeta_+) T.
 \end{aligned} \tag{52}$$

The ratios of variances for vanilla call vegas are shown in Figures 3 and 4. As was the case with the deltas, the AMVD method gives lower variances than the LR method, but, in contrast to the deltas, it also produces lower variances than the MVD method. The AMVD method also produces lower variances than the PW method for in-the-money calls.

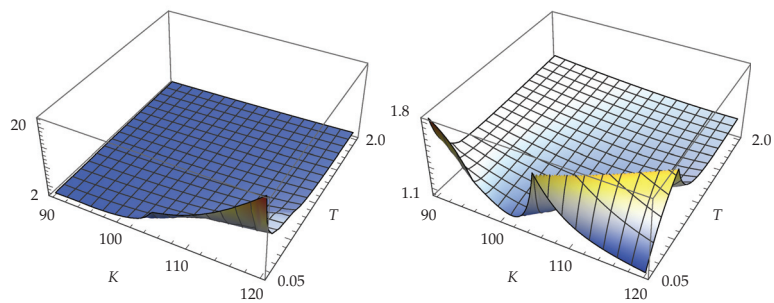


Figure 3. On the left, there is the ratio $\mathbb{V}[\Delta_{\sigma}^{\text{LR}}(c)] / \mathbb{V}[\Delta_{\sigma}^{\text{AMVD}}(c)]$ and, on the right, there is the ratio $\mathbb{V}[\Delta_{\sigma}^{\text{PW}}(c)] / \mathbb{V}[\Delta_{\sigma}^{\text{AMVD}}(c)]$.

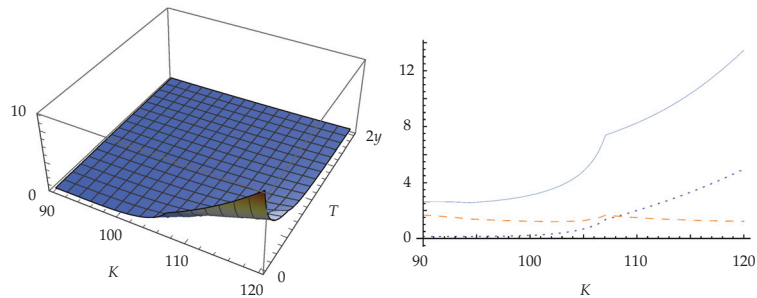


Figure 4. On the left, there is the ratio $\mathbb{V}[\Delta_{\sigma}^{\text{PW}}(c)] / \mathbb{V}[\Delta_{\sigma}^{\text{AMVD}}(c)]$. On the right, there are variance ratios at $T = 0.1$. The solid curve is $\mathbb{V}[\Delta_{\sigma}^{\text{LR}}(c)] / \mathbb{V}[\Delta_{\sigma}^{\text{AMVD}}(c)]$, the dashed curve is $\mathbb{V}[\Delta_{\sigma}^{\text{MVD}}(c)] / \mathbb{V}[\Delta_{\sigma}^{\text{AMVD}}(c)]$, and the dotted curve is $\mathbb{V}[\Delta_{\sigma}^{\text{PW}}(c)] / \mathbb{V}[\Delta_{\sigma}^{\text{AMVD}}(c)]$.

6. Digital Calls

In this section, we compute the LR, MVD, and AMVD variances for the digital call deltas and vegas. Once again, we only provide details for digital calls; let K be the strike for a digital call with expiry T .

6.1. Digital Call Delta

The likelihood ratio delta for the digital call is computed using the expression

$$\mathbb{E}[\Delta_{S_0}^{\text{LR}}(d)] = \int_{-\xi_-}^{\infty} \frac{e^{-rT}\xi}{S_0\zeta} \phi(\xi) d\xi = \frac{e^{-rT}}{S_0\zeta} \phi(\xi_-),$$

and the associated variance is

$$\mathbb{V}[\Delta_{S_0}^{\text{LR}}(d)] = \frac{e^{-2rT}}{S_0^2\zeta^2} [\mathcal{I}(x^2, 0; -\xi_-, \infty) - \phi^2(\xi_-)]. \quad (53)$$

The MVD delta for the digital call is computed using the expression

$$\mathbb{E}[\Delta_{S_0}^{\text{MVD}}(d)] = \int_{0 \vee -\xi_-}^{\infty} \frac{e^{-rT}}{S_0\zeta\sqrt{2\pi}} \sqrt{2\pi} \xi \phi(\xi) d\xi - \int_{0 \wedge -\xi_-}^0 \frac{e^{-rT}}{S_0\zeta\sqrt{2\pi}} \sqrt{2\pi} (-\xi) \phi(\xi) d\xi,$$

and the variance is given by

$$\mathbb{V}[\Delta_{S_0}^{\text{MVD}}(d)] = \frac{e^{-2rT}}{S_0^2\zeta^2\sqrt{2\pi}} [\mathcal{I}(x, 0; 0 \vee -\xi_-, \infty) - \mathcal{I}(x, 0; 0 \wedge -\xi_-, 0) - \sqrt{2\pi}\phi^2(\xi_-)]. \quad (54)$$

Finally, AMVD delta for the digital is given by

$$\mathbb{E}[\Delta_{S_0}^{\text{AMVD}}(d)] = \int_{-\xi_-}^{\infty} \frac{e^{-rT}}{S_0\zeta} \sqrt{\frac{2}{\pi}} \cdot \sqrt{\frac{\pi}{2}} |\xi| \phi(\xi) d\xi,$$

with the associated variance

$$\mathbb{V}[\Delta_{S_0}^{\text{AMVD}}(d)] = \frac{e^{-2rT}}{S_0^2\zeta^2} \sqrt{\frac{2}{\pi}} \left[\mathcal{I}(x, 0; 0 \vee -\xi_-, \infty) - \mathcal{I}(x, 0; 0 \wedge -\xi_-, 0) - \sqrt{\frac{\pi}{2}} \phi^2(\xi_-) \right].$$

We note that the PW method cannot be applied for digital options.

The ratios of variances for digital call delta are shown in Figure 5. The AMVD method gives lower variances than the LR method, particularly for short-dated deep in-the-money digitals. The MVD method gives lower variances than the AMVD method, most noticeably for at-the-money digitals.

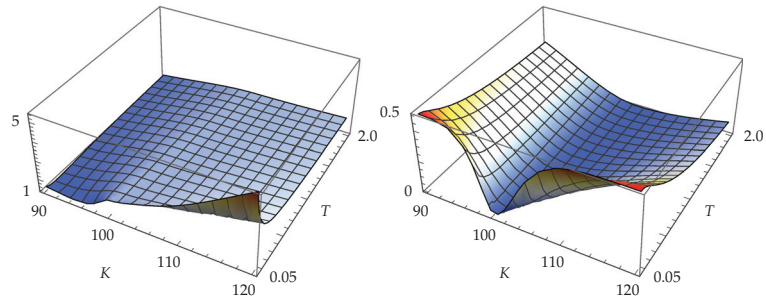


Figure 5. On the left, there is the ratio $V[\Delta_{S_0}^{LR}(d)] / V[\Delta_{S_0}^{AMVD}(d)]$ and, on the right, there is the ratio $V[\Delta_{S_0}^{MVD}(d)] / V[\Delta_{S_0}^{AMVD}(d)]$.

6.2. Digital Call Vega

The likelihood ratio vega for the digital call is computed using the expression

$$\mathbb{E}[\Delta_{\sigma}^{LR}(d)] = \int_{-\xi_-}^{\infty} \frac{e^{-rT}}{\sigma} q_{\xi}(\xi) \cdot \phi(\xi) d\xi = \frac{e^{-rT} \xi_+}{\sigma} \phi(\xi_-),$$

and the associated variance is

$$V[\Delta_{\sigma}^{LR}(d)] = \frac{e^{-2rT}}{\sigma^2} [\mathcal{I}(q_{\xi}^2(x), 0; -\xi_-, \infty) - \xi_+^2 \phi^2(\xi_-)]. \quad (55)$$

The MVD vega for the digital call is computed using the expression

$$\begin{aligned} \mathbb{E}[\Delta_{\sigma}^{MVD}(d)] &= \int_{-\xi_-}^{\infty} \frac{e^{-rT}}{\sigma} \cdot \xi^2 \phi(\xi) d\xi - \int_{0 \vee -\xi_-}^{\infty} \frac{e^{-rT}}{\sigma \sqrt{2\pi}} \cdot \sqrt{2\pi} \xi \phi(\xi) d\xi \\ &+ \int_{0 \wedge -\xi_-}^0 \frac{e^{-rT}}{\sigma \sqrt{2\pi}} \cdot \sqrt{2\pi} (-\xi) \phi(\xi) d\xi - \int_{-\xi_-}^{\infty} \frac{e^{-rT}}{\sigma} \cdot \phi(\xi) d\xi, \end{aligned}$$

and the associated variance is given by

$$\begin{aligned} V[\Delta_{\sigma}^{MVD}(d)] &= \frac{e^{-2rT}}{\sigma^2} [\mathcal{I}(x^2, 0; -\xi_-, \infty) + \mathcal{I}(1, 0; -\xi_-, \infty) - \xi_+^2 \phi^2(\xi_-)] \\ &+ \frac{e^{-2rT}}{\sigma^2 \sqrt{2\pi}} [\mathcal{I}(x, 0; 0 \vee -\xi_-, \infty) - \mathcal{I}(x, 0; 0 \wedge -\xi_-, 0)]. \end{aligned} \quad (56)$$

Finally, the AMVD vega for the digital call is computed using the expression

$$\mathbb{E}[\Delta_{\sigma}^{AMVD}(d)] = \int_{-\xi_-}^{\infty} \frac{e^{-rT} \eta_{\xi}}{\sigma} \cdot \frac{1}{\eta_{\xi}} |q_{\xi}(\xi)| \phi(\xi) d\xi,$$

with the associated variance

$$\begin{aligned} \mathbb{V}[\Delta_{\sigma}^{\text{AMVD}}(d)] = \frac{e^{-2rT}\eta_{\zeta}}{\sigma^2} & \left[\mathcal{I}(q(x), 0; \zeta_+ \vee -\tilde{\zeta}_-, \infty) \right. \\ & - \mathcal{I}(q(x), 0; (\zeta_- \vee -\tilde{\zeta}_-) \wedge \zeta_+, \zeta_+) \\ & \left. + \mathcal{I}(q(x), 0; \zeta_- \wedge -\tilde{\zeta}_-, \zeta_-) - \frac{\zeta_+^2}{\eta_{\zeta}} \phi^2(\tilde{\zeta}_-) \right]. \end{aligned} \quad (57)$$

The ratios of variances for digital calls are shown in Figure 6, where it can be seen that the AMVD method gives lower variance than both the LR and the MVD methods.

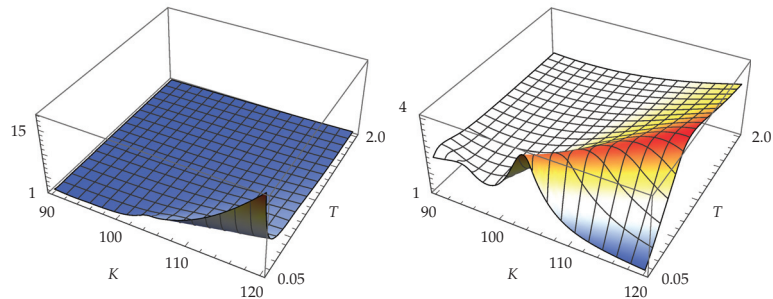


Figure 6. On the left, there is the ratio $\mathbb{V}[\Delta_{\sigma}^{\text{LR}}(d)] / \mathbb{V}[\Delta_{\sigma}^{\text{AMVD}}(d)]$ and, on the right, there is the ratio $\mathbb{V}[\Delta_{\sigma}^{\text{MVD}}(d)] / \mathbb{V}[\Delta_{\sigma}^{\text{AMVD}}(d)]$.

7. Double-Barrier Option

In this section, we consider the delta and vega for a double-barrier option that knocks out if the closing price of the underlying asset lies outside the range, $]B_L, B_U[$, on at least three days prior to expiry, where $B_L < B_U$, and pays the constant rebate $R = 100$ otherwise. We take $S_0 = 100$, $T = 0.25$, $\sigma = 0.20$, and subdivide the interval $[0, 25]$ into 90 equal subintervals. This is the example considered in Chan and Joshi (2015) Subsection 5.3.

7.1. Generation of Non-Normal Random Variates

In order to implement the AMVD method, and the MVD method for comparison purposes, we need to generate random variates from the Rayleigh distribution, the double-sided Maxwell–Boltzmann distribution, and the distributions corresponding to the absolute value densities appearing in (23) and (24).

7.1.1. Rayleigh Random Variate

The standard Rayleigh distribution is defined for $x \geq 0$ and has the probability density function

$$\phi_R(x) = xe^{-\frac{1}{2}x^2}, \quad (58)$$

and the cumulative density function

$$\Phi_R(x) = 1 - e^{-\frac{1}{2}x^2}. \quad (59)$$

Since the inverse of the cumulative density function is given by

$$\Phi_R^{-1}(x) = \sqrt{-2\ln(1-x)} \quad (60)$$

for $x \in [0, 1]$, a variate from the Rayleigh distribution can be generated using the following algorithm:

R1. Generate a uniform variate $u \in [0, 1]$.

R2. Set $x = \Phi_R^{-1}(u) = \sqrt{-2 \ln(1-u)}$, which is then a Rayleigh random variate.

7.1.2. Double-Sided Maxwell–Boltzmann Random Variate

The standard Maxwell–Boltzmann distribution is defined for $x \geq 0$ and has probability density function

$$\phi_{MB}(x) = \sqrt{\frac{2}{\pi}} x^2 e^{-\frac{1}{2}x^2}. \quad (61)$$

An algorithm for generating variates from the Maxwell–Boltzmann distribution based on the acceptance–rejection method, and introduced in Heidergott et al. (2008), is as follows:

MB1. Generate independent uniform variates $u_1, u_2 \in [0, 1]$. Set $w = \sqrt{-4 \ln(u_1)}$.

MB2. If $u_2 > 1.16582199 u_1 w$, then go back to step MB1.³

MB3. Generate a uniform variate $u_3 \in [0, 1]$, independent of u_1 and u_2 .

MB4. If $u_3 < 0.5$, set $x = w$, and set $x = -w$ otherwise. Then, x is a double-sided Maxwell–Boltzmann random variate.

7.1.3. Absolute Rayleigh Distribution

Consider the random variable with domain $\mathbb{R}$, and the probability density function

$$\phi_{|R|}(x) = \frac{1}{2} |x| e^{-\frac{1}{2}x^2}. \quad (62)$$

Note that $\phi_{|R|}$ is obtained by extending ϕ_R to $\mathbb{R}$ using symmetry about the vertical axis. Hence, an algorithm for generating a variate from the *absolute Rayleigh distribution* is as follows:

AR1. Generate a uniform variate $u \in]0, 1[$.

AR2. Then, an absolute Rayleigh random variate, x , is given by

$$x = \begin{cases} \sqrt{-2 \ln(2(1-u))}, & \text{if } u \geq \frac{1}{2}, \\ -\sqrt{-2 \ln(2u)}, & \text{if } u < \frac{1}{2}. \end{cases} \quad (63)$$

7.1.4. Absolute Quadratic Normal (AQN) Distribution

Assume $\varsigma \in \mathbb{R}_+$, and let q_ς , η_ς , and $\varsigma_\pm$ be as defined in (29), (49), and (50), respectively. Consider the random variable with domain $\mathbb{R}$, and the probability density function

$$\phi_{AQN}(x; \varsigma) = \frac{1}{\eta_\varsigma} |q_\varsigma(x)| \phi(x). \quad (64)$$

Note that $\varsigma_\pm$ are the roots of $q_\varsigma(x)$, and $q_\varsigma(x) < 0$ for $x \in]\varsigma_-, \varsigma_+[$. If we denote by Φ_{AQN} the corresponding cumulative density function, then a straight forward calculation gives

$$\Phi_{AQN}(x; \varsigma) = \frac{1}{\eta_\varsigma} \begin{cases} (\varsigma - x) \phi(x), & x \leq \varsigma_-, \\ 2(\varsigma - \varsigma_-) \phi(\varsigma_-) + (x - \varsigma) \phi(x), & \varsigma_- < x \leq \varsigma_+, \\ \eta_\varsigma - (x - \varsigma) \phi(x), & x > \varsigma_+. \end{cases} \quad (65)$$

Note that, in each of the three regions, $\Phi_{AQN}(x; \varsigma)$ has the form

$$\frac{1}{\eta_\varsigma} [\alpha + \epsilon(x - \varsigma) \phi(x)],$$

where $\alpha \in \mathbb{R}$ and $\epsilon = \pm 1$. Thus, given a uniform random variate u , to compute the corresponding variate with density $\phi_{\text{AQN}}(x; \zeta)$, we need to compute the root of a function of the form

$$h(x) = \frac{1}{\eta_{\zeta}} [\alpha + \epsilon(x - \zeta)\phi(x)] - u,$$

with α and ϵ appropriately chosen according to the value of u . For this purpose, we apply Newton's method noting that

$$h'(x) = -\frac{\epsilon}{\eta_{\zeta}} (x^2 - \zeta x - 1)\phi(x) = -\frac{\epsilon}{\eta_{\zeta}} q_{\zeta}(x)\phi(x).$$

Thus, given an initial point $x^{(0)}$, we can apply the recursion

$$x^{(i+1)} = x^{(i)} - \frac{h(x^{(i)})}{h'(x^{(i)})}, \quad (66)$$

until the desired degree of precision is achieved.

However, due to the nature of $h(x)$, Newton's method can be highly sensitive to the initial point $x^{(0)}$. In fact, a poor initial value can result in the method failing to converge to the correct root. Thus, it is crucial that the initial point is chosen carefully, and, for this, we take inspiration from Heidegott et al. (2008) where the Weibull distribution is used in the acceptance–rejection method to generate the double-sided Maxwell–Boltzmann distribution.

For the two regions, $]-\infty, \zeta_-]$ and $]\zeta_+, \infty[$, we approximate the AQN density using Weibull distributions

$$\phi_W^{\pm}(x) = \mp 2\beta_{\pm}(x - \zeta_{\pm})e^{-\beta_{\pm}(x - \zeta_{\pm})^2}, \quad (67)$$

with the cumulative density function

$$\Phi_W^{\pm}(x) = e^{-\beta_{\pm}(x - \zeta_{\pm})^2}. \quad (68)$$

The parameters, $\beta_{\pm}$, are determined by matching the variances of the approximating Weibull distributions to the variances of the AQN variate restricted to the intervals $]-\infty, \zeta_-]$ and $]\zeta_+, \infty[$, respectively. This gives

$$\beta_{\pm} = \frac{1}{\sigma_{\pm}^2} \left(1 - \frac{\pi}{4}\right). \quad (69)$$

where $\sigma_{\pm}^2$ are variances of the AQN variate restricted to $]-\infty, \zeta_-]$ and $]\zeta_+, \infty[$. For the details, refer to Appendix A.

For the region $[\zeta_-, \zeta_+]$, we approximate the AQN cumulative density on $[\zeta_-, \zeta_m]$ and $[\zeta_m, \zeta_+]$, where $\zeta_m = \frac{1}{2}(\zeta_- + \zeta_+)$, using quadratics

$$\begin{aligned} q_-(x) &= c_-(x - \zeta_-)^2, \quad \text{and} \\ q_+(x) &= \Phi_{\text{AQN}}(\zeta_+; \zeta) - \Phi_{\text{AQN}}(\zeta_-; \zeta) - c_+(x - \zeta_+)^2, \end{aligned}$$

with $c_{\pm}$ determined by setting $q_-(\zeta_m) = \Phi_{\text{AQN}}(\zeta_m; \zeta) - \Phi_{\text{AQN}}(\zeta_-; \zeta) = q_+(\zeta_m)$, which gives

$$c_- = \frac{\Phi_{\text{AQN}}(\zeta_m; \zeta) - \Phi_{\text{AQN}}(\zeta_-; \zeta)}{(\zeta_m - \zeta_-)^2}, \quad (70)$$

$$c_+ = \frac{\Phi_{\text{AQN}}(\zeta_+; \zeta) - \Phi_{\text{AQN}}(\zeta_m; \zeta)}{(\zeta_m - \zeta_+)^2}.$$

(71)

The objective function, h , and the initial point, $x^{(0)}$, for the four regions determined by the uniform random variate $u \in]0, 1[$ are shown in Table 1.

Table 1. Objective functions and the initial values for generating the AQN random variate using Newton’s method, where $u_{\pm} = \Phi_{\text{AQN}}(\zeta_{\pm}; \zeta)$.

u	$h(x) + u$	$x^{(0)}$
$]0, u_-]$	$\frac{1}{\eta_{\zeta}}(\zeta - x)\phi(x)$	$\zeta_- - \sqrt{-\frac{1}{\beta_-} \ln\left(\frac{u}{u_-}\right)}$
$]u_-, \frac{1}{2}(u_- + u_+)]$	$\frac{2}{\eta_{\zeta}}(\zeta - \zeta_-)\phi(\zeta_-) + \frac{1}{\eta_{\zeta}}(x - \zeta)\phi(x)$	$\zeta_- + \sqrt{\frac{u - u_-}{c_-}}$
$]\frac{1}{2}(u_- + u_+), u_+]$	$\frac{2}{\eta_{\zeta}}(\zeta - \zeta_-)\phi(\zeta_-) + \frac{1}{\eta_{\zeta}}(x - \zeta)\phi(x)$	$\zeta_+ - \sqrt{\frac{u_+ - u}{c_+}}$
$]u_+, 1[$	$1 - \frac{1}{\eta_{\zeta}}(x - \zeta)\phi(x)$	$\zeta_+ + \sqrt{-\frac{1}{\beta_+} \ln\left(\frac{1 - u}{1 - u_+}\right)}$

Using the initial values given in Table 1, Newton’s method converges within four iterations in most cases. It should be noted that each Newton iteration only involves elementary operations and one exponential, and so the method is quite computationally efficient.

7.2. Double-Barrier Vega

Vegas for the double knock-out barrier was computed using a Monte Carlo simulation implemented in C++ with uniform random variates generated by Mersenne Twister from the boost library.⁴ For each method, 100 independent valuations were performed with each valuation using 50,000 paths for the MVD method and 100,000 paths for the LR and AMVD methods. Standard deviations were then computed from these 100 samples. The lower and upper barriers were set to $L_i = 75 + 2i$ and $U_i = 105 + 2j$ for $0 \leq i, j \leq 10$. The resulting means and standard deviations of vegas for a subset of (L_i, U_i) pairs is shown in Table 2.

Table 2. Means and standard deviations of vegas for double-barrier options with varying lower and upper barriers using the LR, MVD, and AMVD methods.

L	U	LR	Mean		Standard Deviation		
			MVD	AMVD	LR	MVD	AMVD
75	105	−169.34	−167.52	−168.29	15.68	7.69	11.17
75	115	−173.85	−172.96	−171.81	19.51	6.43	13.65
75	125	−60.85	−60.17	−60.39	20.08	3.25	15.7
85	105	−292.55	−291.65	−292.08	12.04	8.6	10.75
85	115	−305.83	−306.81	−307.69	19.13	8.05	14.14
85	125	−193.72	−195.74	−195.11	21.13	6.26	13.96
95	105	−146.66	−147.24	−146.53	5.67	7.77	5.92
95	115	−316.54	−317.12	−318.45	13.47	10.15	10.38
95	125	−220.83	−219.58	−218.19	14.36	7.8	10.75

It is evident from Figure 7 that MVD vegas have the lowest standard deviations in general for double-barrier options, followed by those of the AMVD, and then the LR method. It is worth noting that the standard deviation in vegas for the AMVD method is lowest, in relative terms, when the upper barrier is close to the lower barrier. In fact, the AMVD method has the lowest standard deviation among the three methods when $(L, U) = (95, 105)$.

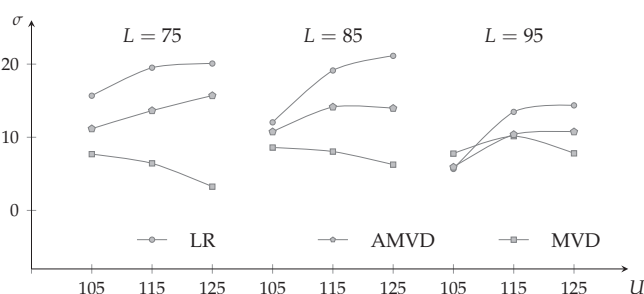


Figure 7. Comparison of double-barrier option vega standard deviations for the LR, MVD, and AMVD methods.

This is consistent with an observation made in previous sections that the method giving sensitivities with the smallest variance will not only depend on the derivative being considered but also on the details such as strikes and barrier levels, and the way each method distributes densities relative to the strikes and barriers. The implementer should experiment and choose the method that is most appropriate for each application.

8. Conclusions

In this paper, we introduced an alternative to the measure-valued differentiation (MVD) method for computing Monte Carlo sensitivities, called the *absolute measure-valued differentiation* (AMVD) method, which decomposes the density derivative as the product of its sign and its absolute value. This method shares many of the properties of the MVD method, but instead of initiating two additional paths in each simulation step of the original path, it initiates only *one* additional subpath. It was shown that the AMVD method gives vega estimates with lower variance than the MVD method for vanilla and digital calls in the Black–Scholes model. Moreover, the AMVD vega estimates have lower variance than the pathwise differentiation method in certain situations, such as the case of short-dated deep in-the-money vanilla calls. A consideration of the vegas for the more exotic double-barrier options revealed that the relative performances of the likelihood ratio, MVD, and AMVD methods depend, in general, on the positions of the barriers, and the way each method distributes densities relative to the barriers.

As is the case with the MVD method, applying the AMVD method requires the generation of nonstandard random variates. For the examples considered in this paper, we provided a highly efficient implementation of the inverse transform method for this purpose using Newton's method with a careful selection of the initial point.

Author Contributions: Conceptualization, M.J. and O.K.K.; methodology, M.J. and O.K.K.; software, O.K.K.; formal analysis, M.J., O.K.K. and S.S.; writing—original draft preparation, M.J. and O.K.K. writing—review and editing, O.K.K. and S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data and software used in this article are available on request from the corresponding author (O.K.).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Initial Values for Newton's Method

For the initial point in Newton's method to invert the AQN cumulative density function, we approximate the cumulative density function using a Weibull density with matching variance. For this purpose, note that the mean, μ_+ , and the second moment, $m_+^{(2)}$, for the AQN distribution over the region $[x_+, \infty[$ are given by

$$\begin{aligned}\mu_+ &= \frac{1}{\eta_+} \int_{\zeta_+}^{\infty} x(x^2 - \zeta_+ x - 1) \phi(x) dx = \frac{1}{\eta_+} \int_{\zeta_+}^{\infty} [(x^3 - 2x) - \zeta_+(x^2 - 1) + x - \zeta_+] \phi(x) dx \\ &= \frac{1}{\eta_+} [(\zeta_+^2 - \zeta_+ \zeta_+ + 1) \phi(\zeta_+) - \zeta_+ \Phi(-\zeta_+)],\end{aligned}$$

where $\eta_+ = (\zeta_+ - \zeta) \phi(\zeta_+) / \eta_{\zeta}$, and

$$\begin{aligned}m_+^{(2)} &= \frac{1}{\eta_+} \int_{\zeta_+}^{\infty} x^2(x^2 - \zeta_+ x - 1) \phi(x) dx \\ &= \frac{1}{\eta_+} \int_{\zeta_+}^{\infty} [(x^4 - 3x^2) - \zeta_+(x^3 - 2x) + 2(x^2 - 1) - 2\zeta_+ x + 2] \phi(x) dx \\ &= \frac{1}{\eta_+} [(\zeta_+^3 - \zeta_+ \zeta_+^2 + 2\zeta_+ - 2\zeta) \phi(\zeta_+) + 2\Phi(-\zeta_+)],\end{aligned}$$

The variance is given by $\sigma_+^2 = m_+^{(2)} - \mu_+^2$, and the corresponding mean and second moment for approximating the Weibull distribution are

$$\begin{aligned}\mu_{W,+} &= 2\beta_+ \int_{x_+}^{\infty} x(x - \zeta_+) e^{-\beta_+(x-\zeta_+)^2} dx = 2\beta_+ \int_0^{\infty} (x + \zeta_+) x e^{-\beta_+ x^2} dx \\ &= \int_0^{\infty} e^{-\beta_+ x^2} dx - \zeta_+ [e^{-\beta_+ x^2}]_0^{\infty} = \frac{1}{2} \sqrt{\frac{\pi}{\beta_+}} + \zeta_+, \end{aligned}$$

and

$$\begin{aligned}m_{W,+}^{(2)} &= 2\beta_+ \int_{\zeta_+}^{\infty} x^2(x - \zeta_+) e^{-\beta_+(x-\zeta_+)^2} dx = 2\beta_+ \int_0^{\infty} (x + \zeta_+)^2 x e^{-\beta_+ x^2} dx \\ &= 2\beta_+ \int_0^{\infty} (x^3 + 2\zeta_+ x^2 + \zeta_+^2 x) e^{-\beta_+ x^2} dx = 2 \int_0^{\infty} [(1 + \beta_+ \zeta_+^2) x + \zeta_+] e^{-\beta_+ x^2} dx \\ &= 2\zeta_+ \int_0^{\infty} e^{-\beta_+ x^2} dx - \frac{(1 + \beta_+ \zeta_+^2)}{\beta_+} [e^{-\beta_+ x^2}]_0^{\infty} = \sqrt{\frac{\pi}{\beta_+}} \zeta_+ + \frac{(1 + \beta_+ \zeta_+^2)}{\beta_+}.\end{aligned}$$

The variance of the approximating the Weibull distribution is given by

$$\begin{aligned}\sigma_{W,+}^2 &= \sqrt{\frac{\pi}{\beta_+}} \zeta_+ + \left(\frac{1}{\beta_+} + \zeta_+^2 \right) - \left[\frac{1}{2} \sqrt{\frac{\pi}{\beta_+}} + \zeta_+ \right]^2 \\ &= \sqrt{\frac{\pi}{\beta_+}} \zeta_+ + \left(\frac{1}{\beta_+} + \zeta_+^2 \right) - \frac{\pi}{4\beta_+} - \sqrt{\frac{\pi}{\beta_+}} \zeta_+ - \zeta_+^2 = \frac{1}{\beta_+} \left(1 - \frac{\pi}{4} \right),\end{aligned}$$

and similar calculations for the region $]-\infty, \zeta_-]$ gives

$$\begin{aligned}\mu_- &= \frac{1}{\eta_-} \int_{-\infty}^{\zeta_-} x(x^2 - \zeta_- x - 1) \phi(x) dx \\ &= \frac{1}{\eta_-} [(-\zeta_-^2 + \zeta_- \zeta_- - 1) \phi(\zeta_-) - \zeta_- \Phi(\zeta_-)], \\ m_-^{(2)} &= \frac{1}{\eta_-} \int_{-\infty}^{\zeta_-} x^2(x^2 - \zeta_- x - 1) \phi(x) dx \\ &= \frac{1}{\eta_-} [(-\zeta_-^3 + \zeta_- \zeta_-^2 - 2\zeta_- + 2\zeta) \phi(\zeta_-) + 2\Phi(\zeta_-)], \\ \sigma_{W,-}^2 &= \frac{1}{\beta_-} \left(1 - \frac{\pi}{4} \right),\end{aligned}$$

where $\eta_- = (\varsigma - \varsigma_-)\phi(\varsigma_-)/\eta_\varsigma$. Equating the corresponding variances gives

$$\beta_{\pm} = \frac{1}{\sigma_{\pm}^2} \left(1 - \frac{\pi}{4}\right). \quad (\text{A1})$$

Appendix B. Vanilla Call Variances

Appendix B.1. Likelihood Ratio Call Delta Variance

The second moment associated with the likelihood ratio delta for the vanilla call is given by

$$\begin{aligned} m^{(2)}[\Delta_{S_0}^{\text{LR}}] &= \int_{-\xi_-}^{\infty} \left[e^{-rT} \left(S_0 e^{(r-\frac{1}{2}\sigma^2)T+\varsigma\zeta} - K \right) \frac{\zeta}{S_0\varsigma} \right]^2 \phi(\zeta) d\zeta \\ &= \frac{e^{-2rT}}{S_0^2\varsigma^2} \int_{-\xi_-}^{\infty} \left[S_0^2 e^{(2r-\sigma^2)T+2\varsigma\zeta} - 2KS_0 e^{(r-\frac{1}{2}\sigma^2)T+\varsigma\zeta} + K^2 \right] \zeta^2 \phi(\zeta) d\zeta. \end{aligned}$$

Computing the three terms, we obtain

$$\begin{aligned} m_{S_0,1}^{\text{LR}} &= \frac{e^{-\varsigma^2}}{\varsigma^2} \int_{-\xi_-}^{\infty} e^{2\varsigma\zeta} \zeta^2 \phi(\zeta) d\zeta = \frac{e^{\varsigma^2}}{\varsigma^2} \mathcal{I}(x^2, 2\varsigma; -\xi_-, \infty), \\ m_{S_0,2}^{\text{LR}} &= -\frac{2Ke^{-(r+\frac{1}{2}\sigma^2)T}}{S_0\varsigma^2} \int_{-\xi_-}^{\infty} e^{\varsigma\zeta} \zeta^2 \phi(\zeta) d\zeta = -\frac{2Ke^{-rT}}{S_0\varsigma^2} \mathcal{I}(x^2, \varsigma; -\xi_-, \infty), \\ m_{S_0,3}^{\text{LR}} &= \frac{K^2 e^{-2rT}}{S_0^2\varsigma^2} \int_{-\xi_-}^{\infty} \zeta^2 \phi(\zeta) d\zeta = \frac{K^2 e^{-2rT}}{S_0^2\varsigma^2} \mathcal{I}(x^2, 0; -\xi_-, \infty). \end{aligned}$$

Appendix B.2. MVD Call Delta Variance

The second moment for the MVD delta estimate is given by

$$\begin{aligned} m^{(2)}[\Delta_{S_0}^{\text{MVD}}] &= \int_{0 \vee -\xi_-}^{\infty} \left[\frac{e^{-rT}}{S_0\varsigma\sqrt{2\pi}} \left(S_0 e^{(r-\frac{1}{2}\sigma^2)T+\varsigma\zeta} - K \right) \right]^2 \cdot \sqrt{2\pi} \zeta \phi(\zeta) d\zeta \\ &\quad + \int_{0 \wedge -\xi_-}^0 \left[\frac{e^{-rT}}{S_0\varsigma\sqrt{2\pi}} \left(S_0 e^{(r-\frac{1}{2}\sigma^2)T+\varsigma\zeta} - K \right) \right]^2 \cdot \sqrt{2\pi} (-\zeta) \phi(\zeta) d\zeta. \end{aligned}$$

Computing the first term, we obtain

$$\frac{e^{-2rT}}{\sqrt{2\pi}S_0^2\varsigma^2} \int_{0 \vee -\xi_-}^{\infty} \left[S_0^2 e^{(2r-\sigma^2)T+2\varsigma\zeta} - 2KS_0 e^{(r-\frac{1}{2}\sigma^2)T+\varsigma\zeta} + K^2 \right] \zeta \phi(\zeta) d\zeta,$$

and the three subterms are given by

$$\begin{aligned} m_{S_0,1}^{\text{MVD}} &= \frac{e^{-\varsigma^2}}{\sqrt{2\pi}\varsigma^2} \int_{0 \vee -\xi_-}^{\infty} e^{2\varsigma\zeta} \zeta \phi(\zeta) d\zeta = \frac{e^{\varsigma^2}}{\sqrt{2\pi}\varsigma^2} \mathcal{I}(x, 2\varsigma; 0 \vee -\xi_-, \infty), \\ m_{S_0,1,2}^{\text{MVD}} &= -\frac{2Ke^{-(r+\frac{1}{2}\sigma^2)T}}{\sqrt{2\pi}S_0\varsigma^2} \int_{0 \vee -\xi_-}^{\infty} e^{\varsigma\zeta} \zeta \phi(\zeta) d\zeta = -\frac{2Ke^{-rT}}{\sqrt{2\pi}S_0\varsigma^2} \mathcal{I}(x, \varsigma; 0 \vee -\xi_-, \infty), \\ m_{S_0,1,3}^{\text{MVD}} &= \frac{K^2 e^{-2rT}}{\sqrt{2\pi}S_0^2\varsigma^2} \int_{0 \vee -\xi_-}^{\infty} \zeta \phi(\zeta) d\zeta = \frac{K^2 e^{-2rT}}{\sqrt{2\pi}S_0^2\varsigma^2} \mathcal{I}(x, 0; 0 \vee -\xi_-, \infty). \end{aligned}$$

Similarly, computing the second term, we obtain

$$\frac{e^{-2rT}}{\sqrt{2\pi}S_0^2\varsigma^2} \int_{0 \wedge -\xi_-}^0 \left[S_0^2 e^{(2r-\sigma^2)T+2\varsigma\zeta} - 2KS_0 e^{(r-\frac{1}{2}\sigma^2)T+\varsigma\zeta} + K^2 \right] (-\zeta) \phi(\zeta) d\zeta,$$

and the three subterms are given by

$$\begin{aligned} m_{S_{0,2,1}}^{\text{MVD}} &= -\frac{e^{-\zeta^2}}{\sqrt{2\pi}\zeta^2} \int_{0 \wedge -\zeta_-}^0 e^{2\zeta\tilde{\zeta}} \tilde{\zeta} \phi(\tilde{\zeta}) d\tilde{\zeta} = -\frac{e^{\zeta^2}}{\sqrt{2\pi}\zeta^2} \mathcal{I}(x, 2\zeta; 0 \wedge -\zeta_-, 0), \\ m_{S_{0,2,2}}^{\text{MVD}} &= \frac{2Ke^{-(r+\frac{1}{2}\sigma^2)T}}{\sqrt{2\pi}S_0\zeta^2} \int_{0 \wedge -\zeta_-}^0 e^{\zeta\tilde{\zeta}} \tilde{\zeta} \phi(\tilde{\zeta}) d\tilde{\zeta} = \frac{2Ke^{-rT}}{\sqrt{2\pi}S_0\zeta^2} \mathcal{I}(x, \zeta; 0 \wedge -\zeta_-, 0), \\ m_{S_{0,2,3}}^{\text{MVD}} &= -\frac{K^2e^{-2rT}}{\sqrt{2\pi}S_0^2\zeta^2} \int_{0 \wedge -\zeta_-}^0 \tilde{\zeta} \phi(\tilde{\zeta}) d\tilde{\zeta} = \frac{K^2e^{-2rT}}{\sqrt{2\pi}S_0^2\zeta^2} \mathcal{I}(x, 0; 0 \wedge -\zeta_-, 0). \end{aligned}$$

Appendix B.3. AMVD Call Delta Variance

The second moment for the AMVD delta estimate is given by

$$\begin{aligned} m^{(2)}[\Delta_{S_0}^{\text{AMVD}}] &= 2 \int_{-\zeta_-}^{\infty} \left[\frac{e^{-rT}}{S_0\zeta\sqrt{2\pi}} \left(S_0e^{(r-\frac{1}{2}\sigma^2)T+\zeta\tilde{\zeta}} - K \right) \right]^2 \cdot \sqrt{2\pi}|\tilde{\zeta}| \phi(\tilde{\zeta}) d\tilde{\zeta} \\ &= 2m^{(2)}[\Delta_{S_0}^{\text{MVD}}]. \end{aligned}$$

Appendix B.4. Likelihood Ratio Call Vega Variance

The second moment associated with the likelihood ratio vega for the vanilla call is given by

$$m^{(2)}[\Delta_{\sigma}^{\text{LR}}] = \frac{e^{-2rT}}{\sigma^2} \int_{-\zeta_-}^{\infty} \left[S_0e^{(r-\frac{1}{2}\sigma^2)T+\sigma\sqrt{T}\tilde{\zeta}} - K \right]^2 q^2(\tilde{\zeta}) \phi(\tilde{\zeta}) d\tilde{\zeta},$$

where $q(x)$ is as defined in (30). Expanding out the first term, we obtain

$$m^{(2)}[\Delta_{\sigma}^{\text{LR}}] = \frac{e^{-2rT}}{\sigma^2} \int_{-\zeta_-}^{\infty} \left[S_0^2e^{(2r-\sigma^2)T+2\zeta\tilde{\zeta}} - 2KS_0e^{(r-\frac{1}{2}\sigma^2)T+\zeta\tilde{\zeta}} + K^2 \right] q^2(\tilde{\zeta}) \phi(\tilde{\zeta}) d\tilde{\zeta}.$$

Computing the three terms gives

$$\begin{aligned} m_{\sigma,1}^{\text{LR}} &= \frac{e^{-\zeta^2}S_0^2}{\sigma^2} \int_{-\zeta_-}^{\infty} e^{2\zeta\tilde{\zeta}} q^2(\tilde{\zeta}) \phi(\tilde{\zeta}) d\tilde{\zeta} = \frac{e^{\zeta^2}S_0^2}{\sigma^2} \mathcal{I}(q^2, 2\zeta; -\zeta_-, \infty), \\ m_{\sigma,2}^{\text{LR}} &= -\frac{2e^{-(r+\frac{1}{2}\sigma^2)T}KS_0}{\sigma^2} \int_{-\zeta_-}^{\infty} e^{\zeta\tilde{\zeta}} q^2(\tilde{\zeta}) \phi(\tilde{\zeta}) d\tilde{\zeta} = -\frac{2e^{-rT}KS_0}{\sigma^2} \mathcal{I}(q^2, \zeta; -\zeta_-, \infty), \\ m_{\sigma,3}^{\text{LR}} &= \frac{e^{-2rT}K^2}{\sigma^2} \int_{-\zeta_-}^{\infty} q^2(\tilde{\zeta}) \phi(\tilde{\zeta}) d\tilde{\zeta} = \frac{e^{-2rT}K^2}{\sigma^2} \mathcal{I}(q^2, 0; -\zeta_-, \infty). \end{aligned}$$

Appendix B.5. MVD Call Vega Variance

The second moment associated with the MVD vega is given by

$$\begin{aligned} m^{(2)}[\Delta_{\sigma}^{\text{MVD}}] &= \frac{e^{-2rT}}{\sigma^2} \int_{-\zeta_-}^{\infty} \left[S_0e^{(r-\frac{1}{2}\sigma^2)T+\zeta\tilde{\zeta}} - K \right]^2 \tilde{\zeta}^2 \phi(\tilde{\zeta}) d\tilde{\zeta} \\ &\quad + \frac{e^{-2rT}T}{2\pi} \int_{0 \vee -\zeta_-}^{\infty} \left[S_0e^{(r-\frac{1}{2}\sigma^2)T+\zeta\tilde{\zeta}} - K \right]^2 \sqrt{2\pi}\tilde{\zeta} \phi(\tilde{\zeta}) d\tilde{\zeta} \\ &\quad + \frac{e^{-2rT}T}{2\pi} \int_{0 \wedge -\zeta_-}^0 \left[S_0e^{(r-\frac{1}{2}\sigma^2)T+\zeta\tilde{\zeta}} - K \right]^2 \sqrt{2\pi}(-\tilde{\zeta}) \phi(\tilde{\zeta}) d\tilde{\zeta} \\ &\quad + \frac{e^{-2rT}}{\sigma^2} \int_{-\zeta_-}^{\infty} \left[S_0e^{(r-\frac{1}{2}\sigma^2)T+\zeta\tilde{\zeta}} - K \right]^2 \phi(\tilde{\zeta}) d\tilde{\zeta}, \end{aligned} \tag{A2}$$

The four terms are as follows:

$$\begin{aligned}
 m_{\sigma,1}^{\text{MVD}} &= \frac{e^{\zeta^2} S_0^2}{\sigma^2} \mathcal{I}(x^2, 2\zeta; -\zeta_-, \infty) - \frac{2e^{-rT} K S_0}{\sigma^2} \mathcal{I}(x^2, \zeta; -\zeta_-, \infty) \\
 &\quad + \frac{e^{-2rT} K^2}{\sigma^2} \mathcal{I}(x^2, 0; -\zeta_-, \infty), \\
 m_{\sigma,2}^{\text{MVD}} &= \frac{e^{\zeta^2} S_0^2 T}{\sqrt{2\pi}} \mathcal{I}(x, 2\zeta; 0 \vee -\zeta_-, \infty) - \frac{2e^{-rT} K S_0 T}{\sqrt{2\pi}} \mathcal{I}(x, \zeta; 0 \vee -\zeta_-, \infty) \\
 &\quad + \frac{e^{-2rT} K^2 T}{\sqrt{2\pi}} \mathcal{I}(x, 0; 0 \vee -\zeta_-, \infty), \\
 m_{\sigma,3}^{\text{MVD}} &= -\frac{e^{\zeta^2} S_0^2 T}{\sqrt{2\pi}} \mathcal{I}(x, 2\zeta; 0 \wedge -\zeta_-, 0) + \frac{2e^{-rT} K S_0 T}{\sqrt{2\pi}} \mathcal{I}(x, \zeta; 0 \wedge -\zeta_-, 0) \\
 &\quad - \frac{e^{-2rT} K^2 T}{\sqrt{2\pi}} \mathcal{I}(x, 0; 0 \wedge -\zeta_-, 0), \\
 m_{\sigma,4}^{\text{MVD}} &= \frac{e^{\zeta^2} S_0^2}{\sigma^2} \mathcal{I}(1, 2\zeta; -\zeta_-, \infty) - \frac{2e^{-rT} K S_0}{\sigma^2} \mathcal{I}(1, \zeta; -\zeta_-, \infty) \\
 &\quad + \frac{e^{-2rT} K^2}{\sigma^2} \mathcal{I}(1, 0; -\zeta_-, \infty).
 \end{aligned}$$

Appendix B.6. AMVD Call Vega Variance

The second moment associated with the AMVD vega is given by

$$\begin{aligned}
 m^{(2)}[\Delta_{\sigma}^{\text{AMVD}}] &= \frac{e^{-2rT} \eta_{\zeta}}{\sigma^2} \int_{\zeta_+ \vee -\zeta_-}^{\infty} \left[S_0 e^{(r-\frac{1}{2}\sigma^2)T + \zeta \zeta} - K \right]^2 q(x) \phi(x) dx \\
 &\quad - \frac{e^{-2rT} \eta_{\zeta}}{\sigma^2} \int_{(\zeta_- \vee -\zeta_-) \wedge \zeta_+}^{\zeta_+} \left[S_0 e^{(r-\frac{1}{2}\sigma^2)T + \zeta \zeta} - K \right]^2 q(x) \phi(x) dx \\
 &\quad + \frac{e^{-2rT} \eta_{\zeta}}{\sigma^2} \int_{\zeta_- \wedge -\zeta_-}^{\zeta_-} \left[S_0 e^{(r-\frac{1}{2}\sigma^2)T + \zeta \zeta} - K \right]^2 q(x) \phi(x) dx.
 \end{aligned} \tag{A3}$$

The three terms are as follows:

$$\begin{aligned}
 m_{\sigma,1}^{\text{AMVD}} &= \frac{e^{\zeta^2} S_0^2 \eta_{\zeta}}{\sigma^2} \mathcal{I}(q^2(x), 2\zeta; \zeta_+ \vee -\zeta_-, \infty) - \frac{2e^{-rT} K S_0 \eta_{\zeta}}{\sigma^2} \mathcal{I}(q^2(x), \zeta; \zeta_+ \vee -\zeta_-, \infty) \\
 &\quad + \frac{e^{-rT} K^2 \eta_{\zeta}}{\sigma^2} \mathcal{I}(q^2(x), 0; \zeta_+ \vee -\zeta_-, \infty), \\
 m_{\sigma,2}^{\text{AMVD}} &= -\frac{e^{\zeta^2} S_0^2 \eta_{\zeta}}{\sigma^2} \mathcal{I}(q^2(x), 2\zeta; (\zeta_- \vee -\zeta_-) \wedge \zeta_+, \zeta_+) \\
 &\quad + \frac{2e^{-rT} K S_0 \eta_{\zeta}}{\sigma^2} \mathcal{I}(q^2(x), \zeta; (\zeta_- \vee -\zeta_-) \wedge \zeta_+, \zeta_+) \\
 &\quad - \frac{e^{-rT} K^2 \eta_{\zeta}}{\sigma^2} \mathcal{I}(q^2(x), 0; \zeta_- \wedge -\zeta_-, \infty), \\
 m_{\sigma,3}^{\text{AMVD}} &= \frac{e^{\zeta^2} S_0^2 \eta_{\zeta}}{\sigma^2} \mathcal{I}(q^2(x), 2\zeta; \zeta_- \wedge -\zeta_-, \zeta_-) - \frac{2e^{-rT} K S_0 \eta_{\zeta}}{\sigma^2} \mathcal{I}(q^2(x), \zeta; \zeta_- \wedge -\zeta_-, \zeta_-) \\
 &\quad + \frac{e^{-rT} K^2 \eta_{\zeta}}{\sigma^2} \mathcal{I}(q^2(x), 0; \zeta_- \wedge -\zeta_-, \zeta_-).
 \end{aligned}$$

Appendix C. Digital Call Variances

Appendix C.1. Likelihood Ratio Digital Call Delta Variance

The second moment associated with the likelihood ratio delta for the digital call is

$$m^{(2)}[\Delta_{S_0}^{\text{LR}}] = \frac{1}{S_0^2 \zeta^2} \int_{-\zeta_-}^{\infty} \zeta^2 \phi(\zeta) d\zeta = \frac{1}{S_0^2 \zeta^2} \mathcal{I}(x^2, 0; -\zeta_-, \infty).$$

Appendix C.2. MVD Digital Call Delta Variance

The second moment associated with the MVD delta for the digital call is

$$\begin{aligned} m^{(2)}[\Delta_{S_0}^{\text{MVD}}] &= \frac{1}{S_0^2 \zeta^2} \int_{0 \vee -\zeta_-}^{\infty} \zeta \phi(\zeta) d\zeta + \frac{1}{S_0^2 \zeta^2} \int_{0 \wedge -\zeta_-}^0 (-\zeta) \phi(\zeta) d\zeta \\ &= \frac{1}{S_0^2 \zeta^2} [\mathcal{I}(x, 0; 0 \vee -\zeta_-, \infty) - \mathcal{I}(x, 0; 0 \wedge -\zeta_-, 0)]. \end{aligned}$$

Appendix C.3. AMVD Digital Call Delta Variance

The second moment associated with the AMVD delta for the digital call is

$$m^{(2)}[\Delta_{S_0}^{\text{AMVD}}] = \frac{2}{S_0^2 \zeta^2} \int_{-\zeta_-}^{\infty} |\zeta| \phi(\zeta) d\zeta = 2m^{(2)}[\Delta_{S_0}^{\text{MVD}}].$$

Appendix C.4. Likelihood Ratio Digital Call Vega Variance

The second moment associated with the likelihood ratio vega for the digital call is

$$m^{(2)}[\Delta_{\sigma}^{\text{LR}}] = \frac{1}{\sigma^2} \int_{-\zeta_-}^{\infty} q^2(\zeta) \phi(\zeta) d\zeta = \frac{1}{\sigma^2} \mathcal{I}(q^2(x), 0; -\zeta_-, \infty).$$

Appendix C.5. MVD Digital Call Vega Variance

The second moment associated with the MVD delta for the digital call is

$$\begin{aligned} m^{(2)}[\Delta_{\sigma}^{\text{MVD}}] &= \frac{1}{\sigma^2} \int_{-\zeta_-}^{\infty} \zeta^2 \phi(\zeta) d\zeta - \frac{1}{\sigma^2} \int_{-\zeta_-}^{\infty} \phi(\zeta) d\zeta \\ &\quad + \frac{1}{\sigma^2} \int_{0 \vee -\zeta_-}^{\infty} \zeta \phi(\zeta) d\zeta - \frac{1}{\sigma^2} \int_{0 \wedge -\zeta_-}^0 (-\zeta) \phi(\zeta) d\zeta \\ &= \frac{1}{\sigma^2} [\mathcal{I}(x^2, 0; -\zeta_-, \infty) - \mathcal{I}(1, 0; -\zeta_-, \infty)] \\ &\quad + \frac{1}{\sigma^2} [\mathcal{I}(x^2, 0; 0 \vee -\zeta_-, \infty) - \mathcal{I}(1, 0; 0 \wedge -\zeta_-, 0)]. \end{aligned}$$

Appendix C.6. AMVD Digital Call Vega Variance

The second moment associated with the AMVD vega for the digital call is

$$\begin{aligned} m^{(2)}[\Delta_{S_0}^{\text{AMVD}}] &= \frac{\eta}{S_0^2 \zeta^2} \int_{\zeta_+ \vee -\zeta_-}^{\infty} q(\zeta) \phi(\zeta) d\zeta - \frac{\eta}{S_0^2 \zeta^2} \int_{(\zeta_- \vee -\zeta_-) \wedge \zeta_+}^{\zeta_+} q(\zeta) \phi(\zeta) d\zeta \\ &\quad + \frac{\eta}{S_0^2 \zeta^2} \int_{\zeta_- \wedge -\zeta_-}^{\zeta_-} q(\zeta) \phi(\zeta) d\zeta \\ &= \frac{\eta}{S_0^2 \zeta^2} [\mathcal{I}(q(x), 0; \zeta_+ \vee -\zeta_-, \infty) - \mathcal{I}(q(x), 0; (\zeta_- \vee -\zeta_-) \wedge \zeta_+, \zeta_+)] \\ &\quad + \frac{\eta}{S_0^2 \zeta^2} \mathcal{I}(q(x), 0; \zeta_- \wedge -\zeta_-, \zeta_-). \end{aligned}$$

Notes

- ¹ This is also the case for the MVD method.
- ² It should be noted, however, that this higher computational burden is partially offset by the MVD sensitivity estimates having lower variance.

- ³ We note that the inequality involving u_2 in Heidergott et al. (2008) is in the opposite direction, which is most likely a typographical error.
- ⁴ Refer to www.boost.org. Accessed 3 June 2022.

References

- Black, Fischer, and Myron Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81: 637–54. [CrossRef]
- Brace, Alan. 2007. *Engineering BGM*. Sydney: Chapman and Hall.
- Capriotti, Luca, and Mike Giles. 2018. Fast Correlation Greeks by Adjoint Algorithmic Differentiation. *arXiv* arXiv:1004.1855v1.
- Chan, Jiun Hong, and Mark Joshi. 2015. Optimal limit methods for computing sensitivities of discontinuous integrals including triggerable derivative securities. *IIE Transactions* 47: 978–97. [CrossRef]
- Glasserman, Paul. 2004. *Monte Carlo Methods in Financial Engineering*. New York: Springer.
- Heidergott, Bernd, Felisa J. Vázquez-Abad, and Warren Volk-Makarewicz. 2008. Sensitivity estimation for Gaussian systems. *European Journal of Operational Research* 187: 193–207. [CrossRef]
- Joshi, Mark S., and Dherminder Kainth. 2004. Rapid and accurate development of prices and Greeks for n -th to default credit swaps in the Li model. *Quantitative Finance* 4: 266–75. [CrossRef]
- Liu, Guangwu, and L. Jeff Hong. 2011. Kernel estimation of the greeks for options with discontinuous payoffs. *Operations Research* 59: 96–108. [CrossRef]
- Lyu, Yuh-Dauh, and Huei-Wen Teng. 2011. Unbiased and efficient Greeks of financial options. *Finance and Stochastics* 15: 141–81. [CrossRef]
- Pflug, Georg Ch, and Philipp Thoma. 2016. Efficient calculation of the Greeks for exponential Lévy processes: An application of measure valued differentiation. *Quantitative Finance* 16: 247–57. [CrossRef]
- Rott, Marius G., and Christian P. Fries. 2005. Fast and Robust Monte Carlo Cdo Sensitivities and Their Efficient Object Oriented Implementation. *SSRN*, May 31.
- Xiang, Jiangming, and Xiaoqun Wang. 2022. Quasi-Monte Carlo simulation for American option sensitivities. *Journal of Computational and Applied Mathematics* 413: 114268. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

The Effect of Short-Sale Restrictions on Corporate Managers

Baixiao Liu ¹, John J. McConnell ² and Andrew Schrowang ^{3,*}

¹ HSBC Business School, Peking University, University Town, Nanshan District, Shenzhen 518055, China; baixiaoliu@phbs.pku.edu.cn

² Krannert Graduate School of Management, Purdue University, 403 West State Street, West Lafayette, IN 47907, USA; mconnj@purdue.edu

³ Department of Finance, College of Business, Florida State University, 821 Academic Drive, Tallahassee, FL 32306, USA

* Correspondence: aschrowang@business.fsu.edu

Abstract: This paper studies the effect of short selling on corporate managers from 2002 through 2010. We examine how the exemption of short-sale uptick tests due to the Regulation SHO pilot program affects managers' decisions to abandon value-reducing acquisition attempts. We find that when deciding whether to abandon value-reducing acquisition attempts during the program, managers of pilot firms, whose stocks are less subject to short-selling impediments, are more sensitive to stock price changes than managers of nonpilot firms. We find no difference in managers' sensitivity prior to nor post SHO. These results indicate that, despite their dislike of short sellers, managers believe that the level of informativeness from capital markets is superior when short sellers are less impeded.

Keywords: short selling; Regulation SHO; corporate acquisitions; managerial learning

JEL Classification: G14; G18; G30; G34

1. Introduction

A maintained virtue of a market economy is that market prices provide valuable information to decision makers and that decision makers take this information into account when making decisions. An important class of decision makers is corporate managers, and an important source of price information for these corporate managers is the firm's stock price. Theoretically, as managers reveal new information about corporate decisions, the stock market reacts, and managers adjust appropriately to this new information. Such a perspective relies on the presumption that market prices are reliable signals that incorporate information reasonably well.

Potentially important to the reliability of market prices are stock short sellers. Short sellers search for information and, when they come upon information that indicates a stock is overpriced, take a position that accords with that view. Empirical studies support the proposition that short sellers, on average, carry out informed trades and managers take in and learn from prices.¹ Nevertheless, the actions of short sellers often arouse the ire of managers whose firms become the targets of short selling. An illustrative example comes from Elon Musk, the CEO of Tesla: "[t]hey're constantly trying to make up false rumors and amplify any negative rumors. It's a really big incentive to lie and attack my integrity".² But that perspective leads to a conundrum: managers rely on price signals, short sellers provide more informative prices, and managers loathe short sellers—do managers ignore the price information provided by short sellers?

In this study, we address that question in the context of corporate takeover attempts when there is a restriction on the role of short sellers. Consider the decision process of an acquiring firm's manager of whether to pursue or abandon a proposed acquisition. Prior to the announcement of the acquisition attempt, the acquiring firm's manager forms their

Citation: Liu, Baixiao, John J. McConnell, and Andrew Schrowang. 2023. The Effect of Short-Sale Restrictions on Corporate Managers. *Journal of Risk and Financial Management* 16: 486. <https://doi.org/10.3390/jrfm16110486>

Academic Editor: Thanasis Stengos

Received: 25 September 2023

Revised: 10 November 2023

Accepted: 13 November 2023

Published: 17 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

expectation of the net present value (henceforth, NPV) of the acquisition. Presumably, managers initiate an acquisition attempt only when the calculated NPV of the acquisition is positive. At the attempt announcement, the stock market reacts, and the price change of the acquiring firm's stock is observed by the manager.

When the stock price change is positive, the manager continues to pursue the acquisition. When the stock price change is negative, the manager may reconsider the opportunity and may choose to discontinue the acquisition effort. In part, the manager's decision depends upon the credibility of the information conveyed by the price signal. If short selling is unimpeded and the manager views short sellers as informed traders, the manager may assign more weight to the price changes than when short selling is impeded. That is, given all else equal, managers may be more likely to discontinue a takeover attempt when the stock price change is more negative and short selling is unimpeded. The alternative, of course, is that managers do not believe that short sellers are more informed and therefore, short sellers, whether constrained or not, do not play a role in the decision-making process of managers.

The setting of our empirical study is a randomized experiment on short selling activities in the US equity markets: the Rule 202T pilot program of Regulation SHO, in which stocks in the Russell 3000 index were ranked by trading volume within each exchange, and every third stock was designated as a pilot stock. From 2 May 2005 to 6 July 2007, pilot stocks were exempt from the tick test for exchange-listed stocks and the bid test for NASDAQ stocks (henceforth, collectively, uptick tests). The exemption from the short-sale uptick tests exogenously eliminated these impediments to short selling in a randomly chosen set of pilot stocks during the program.³ The short-sale uptick tests remained in place for nonpilot stocks during this time period. After 6 July 2007, the SEC eliminated short-sale uptick tests for all exchange-listed stocks.

This setting gives rise to three time periods over which we conduct our tests. The first we label as the pre-SHO period, the second as the SHO period, and the third as the post-SHO period. The first period encompasses 1 January 2002–1 May 2005; the second encompasses 2 May 2005–6 August 2007; and the third encompasses 7 August 2007–31 December 2010. During the pre-SHO period, short sellers in all stocks were subject to the uptick tests. During the SHO period, only short sellers in nonpilot stocks were subject to the uptick tests. During the post-SHO period, short sellers in no stocks were subject to the uptick tests. This setting allows for a comparison of the difference between managers of pilot and nonpilot firms in their sensitivity to stock price changes in making the decision to abandon (or complete) their proposed acquisitions when short sellers are impeded and unimpeded by the uptick tests in participating in the price setting process.

We examine the question in a linear probability regression analysis using 756 proposed acquisitions, each with a transaction value of at least USD 100 million and each of which was accompanied by a negative stock price reaction at its announcement during the period of 1 January 2002 and 31 December 2010. The dependent variable in our analysis is whether the proposed acquisition is abandoned (or not). The key independent variables are the acquiring firm's stock price changes at the announcement of the acquisition attempt (i.e., the acquiring firm's 3-day announcement period CAR), an indicator that is assigned the value of one if the acquiring firm's stock is a pilot stock, an indicator that is assigned the value of one if the acquisition is announced during the SHO period, and, importantly, the interaction of these three variables.

To begin, we find that proposed value-reducing acquisitions⁴ are more likely to be abandoned if the stock price reaction is more negative at the announcement of the proposed transaction. That is, we find, as do others, that managers are sensitive to the stock price reaction in making the decision to abandon value-reducing acquisition attempts.⁵

We further find that, during the pre-SHO period, the sensitivity to stock price changes in making the decision to abandon value-reducing acquisition attempts is not different between managers of pilot and nonpilot firms. That is, during this period in which the short selling of all stocks was impeded by the uptick tests, managers of pilot and nonpilot

firms are equally sensitive to stock price reactions in assessing whether to abandon value-reducing acquisition attempts.

In contrast, during the SHO period, in making that decision, managers of pilot firms are more sensitive to stock price changes than are managers of nonpilot firms. In particular, the coefficient of the triple interaction term is negative and statistically significant with a *p*-value less than 0.01. The significance of this variable indicates that managers are more sensitive to stock price changes when short selling is less impeded. To be more emphatic, in making potentially critical decisions about major acquisition attempts, managers are more sensitive to the stock price reaction when short sellers are more likely to be playing a role in the price setting process. Considering that the average likelihood of the abandonment of value-reducing acquisition attempts during the full sample period is 4.23% and, holding constant the acquiring firm's announcement period CAR at its mean of −3.81%, managers of pilot firms are 7.95% more likely to abandon the proposed acquisitions than nonpilot firms during the SHO period compared to the pre-SHO period.

In sum, despite managers' dislike for short sellers, they are more, rather than less, sensitive to price changes when short sellers are less impeded in playing a role in the price setting process. This relation provides evidence that short-selling restrictions could have significant implications for managers in their decision-making process.

In addition, we find that managers of pilot firms are more likely to abandon value-reducing acquisition attempts during the SHO period than the managers of nonpilot firms when the target is in an industry that is different from the acquirer's industry or when the target is a high-tech firm. We further find that managers of pilot firms with stronger shareholder rights are more sensitive to the stock price reaction in making the decision to abandon value-reducing acquisition attempts during the SHO period than the managers of nonpilot firms.⁶

In all, our further examinations of managers' sensitivity to the stock price reaction in the decision to abandon (or complete) value-reducing acquisition attempts based either on the type of target or on the corporate governance of the acquirer indicate that when short sellers are less impeded in the price setting process, managers are more sensitive to stock price reactions in their decisions to abandon (or complete) an acquisition when the acquirer's operations are different from those of the target or when the acquirer is guided by stronger corporate governance.

To put our study in context, we view our investigation as being related to two lines of prior research. Firstly, our study relates to the literature on the corporate governance role of short sellers. This literature includes Massa et al. (2015), Fang et al. (2016), He et al. (2023), and Chang et al. (2019). These studies consider managers' perception of the latent threat of potential short sellers and whether that perception of that threat influences managerial actions. Our study focuses on managers' perception of the manifest effect of short sellers on stock price and whether that perception of manifest short sales influences managerial actions; that is, are managers more likely to learn from stock price movements when short sellers are more likely to play a role in the price setting process? We find that they are.

Secondly, our study relates to a set of studies that investigate the ways in which managers take actions to obstruct short selling of their firm's shares. This literature includes Lamont (2012), Ljungqvist and Qian (2016), and Engelberg et al. (2018). In contrast, our study provides evidence, perhaps the first, that managers, despite their professed distrust of short sellers, are sensitive to the price information provided by short sellers in making corporate decisions.

The remainder of the paper is organized as follows. The next section sets forth the hypothesis to be tested. Section 3 describes the sample and provides the sources of the data used in the empirical analysis. Section 4 presents the primary empirical results. Section 5 presents robustness tests. Section 6 addresses the concerns raised by Litvak et al. (2020) and Heath et al. (2023) about the use of SHO as a setting for quasi-natural experiment of short selling activities. Section 7 summarizes and concludes the study.

2. Hypothesis Development

Prior studies report that a manager's decision to abandon a proposed corporate acquisition is negatively correlated with the stock price reaction at the announcement of the proposed acquisition (Luo 2005; Chen et al. 2007; Kau et al. 2008; Masulis et al. 2009). The more negative the stock price reaction, the more likely the proposed acquisition is to be abandoned. The common interpretation of this result is that managers "listen to the market" when deciding whether to abandon (or complete) their proposed acquisitions. The question that we address here is whether managers are more (or less) likely to listen to the market when short sellers are less impeded in playing a role in the price setting process.

On the one hand, the sentiments expressed by Elon Musk and similarly inclined CEOs lead to the conclusion that "lying" short sellers are not to be trusted. It is a short step from there to the implication that stock prices set, in part, by the unimpeded trading of short sellers are less informative than when short sellers' trading activities are impeded, and from there it is a short step to the implication that managers should ignore the stock price reaction to an acquisition announcement when short sellers are unimpeded in their trading activities.

On the other hand, a review of the literature on the trading activities of short sellers leads to the conclusion that short sellers are astute traders who understand the price setting process well.⁷ It is a short step from there to the implication that stock prices set in part by the unimpeded trading of short sellers are more informative than when short sellers' trading activities are impeded, and from there it is a short step to the implication that managers should be especially sensitive to the stock price reaction to an acquisition announcement when short sellers are unimpeded in their trading activities.

Using the papers discussed above, we developed two hypotheses to explain whether managers listen to the market more when short sellers are less impeded in playing a role in the price setting process:

H₀. *Corporate managers do not value the role short sellers play in the price setting process; therefore, managers will not learn more from the market when short sellers are more or less impeded in the market;*

H_A. *Corporate managers believe that short sellers, on average, do make informed trades; therefore, managers take in and learn from these more informed prices when short sellers are less impeded in playing a role in the price setting process. This leads to managers being more likely to listen to and learn from the market when making corporate decisions, including abandoning a value-reducing acquisition attempt.*

We use the Regulation SHO pilot program to examine these alternative possibilities. The pilot program classifies firms into two groups, pilot and nonpilot firms, wherein pilot firms were (exogenously) selected to be exempt from short-sale uptick tests during the program. This exemption allows for short sellers' involvement in the price setting process to be less impeded. We use this difference to test whether managers are more (or less) sensitive to stock price changes when short sellers are less impeded.

3. Data and Variable Construction

3.1. Value-Reducing Acquisition Attempts

Using the list published by the SEC, we identify an initial sample of 986 pilot stocks and 1966 nonpilot stocks.⁸ Of the 986 pilot stocks, 49.9% (492) are listed on NYSE, 47.9% (472) are listed on NASDAQ-NM, and 2.2% (22) are listed on Amex. The distribution of the nonpilot stocks is similar with 50.0% (982) listed on NYSE, 48.0% (944) on NASDAQ-NM, and 2.0% (40) on Amex.

We obtain proposed acquisitions from the *Thomson Financial Securities Data Company's (SDC) US Mergers and Acquisitions* database. We begin with all transactions classified as a "merger", "acquisition", or "acquisition of a majority interest" that were announced between 1 January 2002 and 31 December 2010.

To be included in the sample for analysis (1) the potential acquirer must be a publicly traded US firm with stock price data available on the *Center for Research in Security Prices* (CRSP) database during the period of 53 trading days prior to the announcement through one trading day after the announcement of the acquisition; (2) the firm must have accounting data available on the *Compustat* database in the year prior to the announcement of the acquisition attempt; (3) the acquirer must be designated as a pilot or nonpilot firm at the start of the Regulation SHO pilot program; (4) the proposed acquirer must own less than 50% of the target firm’s shares prior to the announcement of the acquisition attempt and must seek to own 100% of the target firm’s shares as a result of the acquisition; (5) the proposed transaction must have a value of at least USD 100 million; and (6) the acquisition attempt must be classified as “withdrawn” or “completed”. Table 1 reports the number of observations dropped due to each of these requirements.

Table 1. Sample Selection Procedure. This table presents the sample selection procedure of acquisition attempts obtained from the *Thomson Financial SDC Mergers and Acquisitions* database. The pre-SHO column includes corporate acquisition attempts announced by pilot and nonpilot firms during 1 January 2002 and 1 May 2005. The SHO column includes corporate acquisition attempts announced by pilot and nonpilot firms during 2 May 2005 and 6 August 2007. The post-SHO column includes corporate acquisition attempts announced by pilot and nonpilot firms during 7 August 2007 and 31 December 2010.

	Number of Observations		
	Pre-SHO	SHO	Post-SHO
All acquisition attempts in <i>SDC</i>	5610	4419	4643
Less: missing data in <i>SDC</i> ⁹	2428	1916	2331
Less: missing data in <i>CRSP</i>	15	6	8
Less: missing data in <i>Compustat</i>	279	247	211
Less: acquirer is not designated as a pilot or nonpilot firm	891	819	990
Less: acquirer owns more than 50% of target at announcement	29	13	13
Less: acquirer seeks to own less than 100% of target	164	89	95
Less: transaction value less than USD 100 million	1232	775	539
Less: acquisition attempt is not value-reducing	293	295	238
Value-reducing acquisition attempts	279	259	218

These criteria produce a total of 1582 acquisition announcements by 801 potential acquiring firms over the nine-year period of 2002–2010.¹⁰ For the purposes of this study, we focus on acquisition attempts that are accompanied by a negative stock price reaction at their announcement (“value-reducing acquisition attempts”). To identify such attempts, we compute the cumulative abnormal return (henceforth, CAR) using the market model. We compute CARs using a three-day interval (−1, +1), where event day 0 is the announcement day of the proposed acquisition. Abnormal returns are calculated as the residuals from a market model with the maximum estimation interval of (−252, −10) trading days and the minimum interval of (−52, −10) trading days. The market return used in the estimation is the CRSP value-weighted return. Focusing only on value-reducing acquisition attempts results in a set of 756 acquisition announcements by 473 acquirers. There are 279 acquisition announcements in the *pre-SHO* sample, 259 in the *SHO* sample, and 218 in the *post-SHO* sample period. In terms of the number of observations per year, this sample size is similar to prior studies that focus on value-reducing acquisitions.¹¹

3.2. Key Independent Variables

We create an indicator variable *Pilot* to denote stocks that were selected as pilot stocks in the Regulation SHO pilot program. Pilot stocks are the treatment sample, and nonpilot stocks are the control sample. We construct three variables to indicate three subperiods: *pre-SHO* equals one if the acquisition attempt announcement falls between 1 January 2002 and 1 May 2005 and zero otherwise; *SHO* equals one if the acquisition attempt announcement falls between 2 May 2005 and 6 August 2007 and zero otherwise; and *post-SHO* equals one if the acquisition attempt announcement falls between 7 August 2007 and 31 December 2010 and zero otherwise.

3.3. Control Variables

To isolate the effect of short selling on the acquirer's decision to abandon (or complete) an acquisition that has been announced, we control for other variables that have been shown in prior studies to be correlated with the likelihood of acquisition abandonments. Walkling (1985) and Kau et al. (2008) find that when the target firm undertakes defensive tactics to fend off an unwanted acquisition attempt (*Defense Dummy*), the transaction is more likely to be abandoned. Walkling (1985), Jennings and Mazzeo (1991), Kau et al. (2008), and Masulis et al. (2009) report that the emergence of a competing acquirer (*Compete Dummy*) is positively correlated with the likelihood of transaction abandonment. Huang and Walkling (1987) and Kau et al. (2008) find that when the proposed method of payment includes the stock of the acquiring firm (*Stock Dummy*), the transaction is more likely to be abandoned. Bates and Lemmon (2003) and Kau et al. (2008) report that a receptive "attitude" on the part of the target (*Hostile Dummy*), the presence of tender offers (*Tender Offer Dummy*), and the presence of termination fees (*Termination Fee Dummy*) all decrease the likelihood of abandonment. Burch (2001) and Kau et al. (2008) report that when the proposed transaction includes an option for the acquirer to purchase shares at a fixed price even if a competing offer emerges (*Lockup Dummy*), the acquisition attempt is less likely to be abandoned. Chen et al. (2007) and Kau et al. (2008) find that the larger the market capitalization of the equity of the potential acquirer ($\ln(\text{Market Cap})$), the less likely the transaction is to be abandoned. Luo (2005) finds that if the acquiring and target firm are headquartered in the same state (*Same State HQ Dummy*), the acquisition attempt is less likely to be abandoned, and if the acquisition has a definitive agreement (*Definitive Agreement Dummy*), the acquisition attempt is more likely to be abandoned.

3.4. Descriptive Statistics

Table 2 presents the time series and industry composition of the acquisition attempts according to the primary industry of the acquiring firm. Panel A reports the distribution of acquisition announcements across years; Panel B shows the distribution across industries. This sample covers all 11 of the Global Industry Classification Standard (GICS) sectors.

Table 3 presents descriptive statistics of the independent variables for the pilot and nonpilot firms along with statistical tests that show whether the means and medians are different between the two groups. As shown in Panels A and B, we do find some fundamental differences between the two types of acquisitions. Pilot firms are smaller and have a larger equity Market-to-Book (M/B) ratio than nonpilot firms; the targets of pilot firms are more likely to have multiple potential acquirers; and when the acquirer is a pilot firm, the acquirer is more likely to be headquartered in the same state as the target.

Table 2. Distribution of value-reducing acquisition attempts by SHO program across years and industries. This table presents the distribution of the sample of value-reducing acquisition attempts across years in Panel A and across industries in Panel B for Regulation SHO pilot and nonpilot acquirers over the period of 1 January 2002 through 31 December 2010, obtained from the *Thomson Financial SDC Mergers and Acquisitions* database.

Panel A: Distribution of acquisition attempts across years				
Year	Pilot Acquirers	Nonpilot Acquirers	% Pilot Acquirers	
2002	15	39	27.78%	
2003	21	49	30.00%	
2004	46	74	38.33%	
2005	39	78	33.33%	
2006	46	65	41.44%	
2007	33	65	33.67%	
2008	24	48	33.33%	
2009	14	25	35.90%	
2010	31	44	41.33%	
Total	269	487	35.58%	
Panel B: Distribution of acquisition attempts across industries				
Industry	N	% of Acquisitions Attempts	Pilot Acquirers	% Pilot Acquirers Acquisitions
Energy	65	8.60%	31	32.29%
Materials	23	3.04%	8	25.81%
Industrials	73	9.66%	23	23.96%
Consumer Discretionary	42	5.56%	16	27.59%
Consumer Staples	17	2.25%	8	32.00%
Health Care	114	15.08%	38	25.00%
Financials	139	18.39%	54	27.98%
Information Technology	167	22.09%	54	24.43%
Telecommunication Services	28	3.70%	7	20.00%
Utilities	19	2.51%	8	29.63%
Real Estate	69	9.13%	22	24.18%
Total	756	100.00%	269	35.58%

Table 3. Descriptive statistics. This table presents descriptive statistics for value-reducing acquisition attempts by the Regulation SHO pilot and nonpilot firms obtained from the *Thomson Financial SDC Mergers and Acquisitions* database over the period of 1 January 2002 through 31 December 2010. Panels A and B show the means and medians for acquirers and transaction characteristics, respectively, both for the full sample and for subsamples of only pilot and only nonpilot acquirers. All variables are defined in Appendix A. Statistical tests for differences in means and medians for each characteristic for pilot and nonpilot acquirers are also presented. ***, **, and * indicate significance at 1%, 5%, and 10%, respectively.

	Full Sample		Pilot Acquirers		Nonpilot Acquirers		Difference	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Panel A. Acquirer and acquisition characteristics								
Acquirer Size	17,477.24	3396.49	13,173.03	3334.04	19,854.72	3430.69	−6681.69 **	−96.65
Equity Market-to-Book Ratio	2.58	2.41	3.24	2.42	2.21	2.40	1.03 **	0.02
Log (Firm Age)	2.94	3.00	2.91	2.94	2.95	3.00	−0.05	−0.05
Transaction Value	1785.32	289.20	2104.46	300.00	1609.04	275.00	495.42	25.00
CAR (%)	−3.81	−2.22	−4.14	−2.59	−3.63	−2.13	−0.52	−0.46 *

Table 3. Cont.

	Full Sample		Pilot Acquirers		Nonpilot Acquirers		Difference	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Panel B: Percentage of acquisition attempts with the following characteristics (in %)								
Defense Dummy	1.06	-	1.49	-	0.82	-	0.67	-
Compete Dummy	2.51	-	4.09	-	1.64	-	2.45 *	-
Hostile Dummy	0.79	-	0.74	-	0.82	-	−0.08	-
Tender Offer Dummy	3.57	-	4.09	-	3.29	-	0.80	-
Lockup Dummy	0.53	-	0.74	-	0.41	-	0.33	-
Termination Fee Dummy	31.08	-	31.60	-	30.80	-	0.80	-
Stock Dummy	62.43	-	65.06	-	60.99	-	4.07	-
Same State HQ Dummy	26.46	-	30.11	-	24.44	-	5.68 *	-
Definitive Agreement Dummy	95.37	-	94.05	-	96.10	-	−2.05	-

4. Empirical Results

In this section, we compare the difference between the managers of pilot and nonpilot firms in their sensitivity to stock price changes in making the decision to abandon acquisition attempts when short sellers are impeded and unimpeded by the uptick tests in the price setting process.

4.1. Univariate Analysis of Abandonments of Corporate Acquisition Attempts

We first compare the percentage of proposed value-reducing acquisition attempts abandoned by pilot and nonpilot firms during the *pre-SHO*, *SHO*, and *post-SHO* period. The comparison is illustrated in Figure 1. During the *pre-SHO* period, the difference between pilot and nonpilot acquirers in terms of the percentage of value-reducing acquisition attempts abandoned was −0.086%. This difference between pilot and nonpilot acquirers in abandonment rate increases to 1.329% during the *SHO* period. During the *post-SHO* period, the difference in abandonment rate reverts to −0.072%.

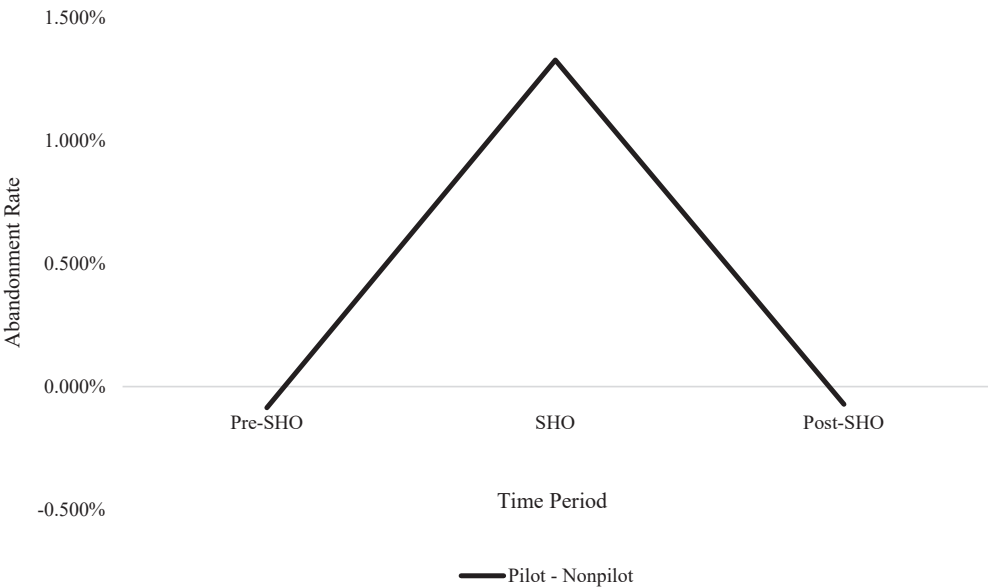


Figure 1. Differences in abandonment rates of value-reducing acquisition attempts between pilot and nonpilot firms during the *pre-SHO*, *SHO* and *post-SHO* Periods. This figure depicts the differences in

abandonment rates between Regulation SHO pilot and nonpilot firms of value-reducing acquisition attempts during the pre-SHO, SHO and post-SHO periods of the Regulation SHO pilot program. The pre-SHO period is 1 January 2002 through 1 May 2005. The SHO period is 2 May 2005 through 6 August 2007. The post-SHO period is 7 August 2007 through 31 December 2010. The sample is from the 2004 Russell 3000 Index as of June 2002. The difference in abandonment rates is calculated as the difference between the abandonment rate for pilot acquirers and the abandonment rate for nonpilot acquirers of proposed value-reducing acquisitions during the time period.

Consistent with the alternative hypothesis, these univariate results support the proposition that, during the pre-SHO period, when short selling is equally impeded by the uptick rule, managers of pilot and nonpilot firms are no different in their propensity to abandon proposed value-reducing acquisitions. In contrast, during the SHO period, managers of pilot firms are more likely to abandon proposed value-reducing acquisitions than managers of nonpilot firms when pilot firms are unimpeded by the uptick rule. Further, during the post-SHO period, when short selling is equally unimpeded by the uptick rule, managers of pilot and nonpilot firms are no different in their propensity to abandon proposed value-reducing acquisitions.

4.2. Linear Probability Regressions of Acquisition Attempt Abandonments

4.2.1. Acquiring Firm’s CAR

We now examine the relation between the likelihood of abandonment of value-reducing acquisition attempts and the acquiring firm’s stock price reaction at the announcement of the attempt controlling for other factors. To do so, we estimate the following linear probability model:

$$Abandonment = \beta_0 + \beta_1 CAR + \gamma Controls + \delta YearDummies + \theta IndustryDummies + \varepsilon, \tag{1}$$

where the dependent variable equals one for abandoned acquisition attempts and zero otherwise. The coefficient of interest is β_1 .

The results of the regression are reported in Column 1 of Table 4. The estimated coefficient of CAR is -0.582 with a p-value less than 0.01; that is, the more negative the stock price reaction, the greater the likelihood that the proposed acquisition will be abandoned across the full sample of value-reducing acquisition attempts. This result is consistent with the findings of prior studies that managers are sensitive to stock price changes when deciding whether to abandon (or complete) proposed value-reducing acquisitions.¹²

Table 4. Acquisition abandonment during the pre-SHO, SHO, and post-SHO periods. This table presents the estimates of a triple difference analysis in a linear probability model of acquisition abandonment. The sample includes 756 value-reducing acquisition attempts by the Regulation SHO pilot and nonpilot firms announced over the period of 1 January 2002 through 31 December 2010. The dependent variable is an indicator that is given the value of one for abandoned acquisition attempts and zero for completed attempts. An acquisition attempt is classified as pilot if the acquirer was designated as a pilot firm during the Regulation SHO pilot program and is designated as SHO or post-SHO based on the announcement date of the acquisition attempt. Variables are defined in Appendix A. All regressions control for year and industry fixed effects. The coefficient estimates of the constant, year, and industry dummies are omitted for brevity. Heteroskedasticity-robust standard errors are clustered by industry, and t-statistics are reported in brackets. ***, **, and * indicate significance at 1%, 5%, and 10%, respectively.

	Dependent Variable: Abandonment (1,0)	
	(1)	(2)
CAR × Pilot × SHO		−2.087 *** [−2.67]
CAR × Pilot × Post-SHO		−0.910 [−0.82]

Table 4. Cont.

	Dependent Variable: Abandonment (1,0)	
	(1)	(2)
CAR	−0.582 *** [−2.71]	−1.645 *** [−5.53]
Pilot		−0.001 [−0.05]
SHO		0.032 [1.60]
Post-SHO		0.008 [0.33]
CAR × Pilot		0.720 [0.87]
CAR × SHO		1.850 *** [4.96]
CAR × Post-SHO		1.820 *** [5.00]
Pilot × SHO		−0.041 [−1.60]
Pilot × Post-SHO		0.006 [0.13]
Defense Dummy	0.353 *** [2.76]	0.371 *** [3.25]
Compete Dummy	0.334 *** [4.64]	0.342 *** [4.35]
Hostile Dummy	0.518 ** [2.22]	0.533 ** [2.40]
Tender Offer Dummy	−0.029 [−0.38]	−0.030 [−0.47]
Lockup Dummy	−0.129 [−0.49]	−0.164 [−0.68]
Termination Fee Dummy	0.004 [0.42]	0.004 [0.44]
Stock Dummy	0.010 [0.92]	0.009 [0.89]
Ln(Market Cap)	−0.006 [−1.08]	−0.006 [−1.13]
Same State HQ Dummy	−0.023 *** [−3.13]	−0.030 *** [−3.72]
Definitive Agreement Dummy	0.074 *** [3.88]	0.070 *** [3.46]
Industry Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes
Observations	756	756
Adj. R ²	0.2397	0.2699

4.2.2. Acquiring Firm’s CAR and Regulation SHO

We next examine whether acquiring firm managers’ sensitivity to stock price reaction in the decision to abandon proposed value-reducing acquisitions is greater (or less) when short selling of the acquiring firm’s stocks is unimpeded by the uptick rule in a multivariate setup. Specifically, we estimate the following linear probability model:

$$\begin{aligned} Abandonment = & \beta_0 + \beta_1 CAR + \beta_2 Pilot + \beta_3 SHO + \beta_4 PostSHO + \beta_5 CAR \times Pilot + \\ & \beta_6 CAR \times SHO + \beta_7 CAR \times PostSHO + \beta_8 Pilot \times SHO + \beta_9 Pilot \times PostSHO + \\ & \beta_{10} CAR \times Pilot \times SHO + \beta_{11} CAR \times Pilot \times PostSHO + \gamma \text{ Control Variables} + \\ & \delta \text{ Year Dummies} + \theta \text{ Industry Dummies} + \epsilon, \end{aligned}$$

(2)

where the dependent variable equals one for abandoned acquisition attempts and zero otherwise. The coefficient estimates of interest are those of the two triple interaction terms, β_{10} for the variable $CAR \times Pilot \times SHO$ and β_{11} for the variable $CAR \times Pilot \times PostSHO$.

The results of the multivariate analysis are given in Column 2 of Table 4. Several items merit comment. Firstly, the coefficient estimate of *Pilot*, β_2 , is -0.001 with a p -value of 0.96 , indicating that, during the *pre-SHO* period, when short selling is equally impeded by the uptick rule for both pilot and nonpilot firms, the sensitivity of managers to stock price changes when deciding whether to abandon value-reducing acquisition attempts is no different between pilot and nonpilot firms.

Secondly, the estimated coefficient of the triple interaction term, $CAR \times Pilot \times SHO$, is -2.087 with a p -value < 0.01 , indicating that holding CAR constant at the mean of -3.81% , managers of pilot acquiring firms are 7.95% more likely to abandon proposed value-reducing acquisitions than are managers of nonpilot acquiring firms during the *SHO* period compared to the *pre-SHO* period. Considering that the average likelihood of abandonment of a value-reducing acquisition attempt during our sample period is 4.23% , a 7.95% increase translates into a 187.94% increase in the likelihood of abandoning a proposed value-reducing acquisition during the *SHO* period compared to the *pre-SHO* period. Consistent with the alternative hypothesis, these findings indicate that, during the *SHO* period, when short selling is unimpeded by the uptick rule for pilot firms, managers of pilot firms, in comparison with managers of nonpilot firms, are more sensitive to stock price changes in the decision to abandon proposed value-reducing acquisitions. Furthermore, this provides evidence that even a single short-sale constraint could have significant negative implications on overall market conditions and the informativeness of the overall market for corporate managers.

Third, the coefficient estimate of $CAR \times Pilot \times PostSHO$, β_{11} , is -0.910 with a p -value of 0.41 . This result shows that, during the *post-SHO* period, when short selling is equally unimpeded by the uptick rule for both pilot and nonpilot firms, the sensitivity of managers to stock price changes in making the decision to abandon value-reducing acquisition attempts is no different between managers of pilot and nonpilot firms.

As for the control variables, acquisition attempts that are classified as hostile attempts, attempts where a competing acquirer emerges, attempts that have a definitive agreement, or attempts in which the target firm undertakes defensive tactics to fend off an unwanted acquisition attempt are more likely to be abandoned. Acquisition attempts in which the target and acquiring firms are headquartered in the same state are less likely to be abandoned. These findings are all statistically significant with p -values < 0.10 .

4.3. Acquisition Characteristics

In this section, we examine whether managers are more sensitive to the stock price reaction when short selling is less impeded by regulation and when managers are likely to put more weight on market prices in their decisions to abandon proposed acquisitions. In particular, we split the sample based on variables that have been shown to proxy for the opacity of an acquisition.¹³

4.3.1. Focused vs. Diversifying Acquisitions

We first split the sample based on the type of acquisition. We classify observations into two samples: diversifying and focused acquisition attempts. We define *diversifying acquisition attempts* as those in which the target and acquiring firms are not in the same Fama–French 17 industry. *Focused acquisition attempts* are defined as those in which the target, and acquiring firms are in the same Fama–French 17 industry. Aggarwal and Baxamusa (2013) have shown that *diversifying acquisition attempts* are positively correlated with multiple measures of information asymmetry, and thus, we use *diversifying acquisition attempts* as a measure of the opacity of the target to the acquirer. We conjecture that managers are likely to put more weight on market prices when the acquisition is a diversifying acquisition.

We re-estimate the model in Equation (2) for each sample. The results are reported in Table 5. Column 1 includes all *diversifying acquisition attempts*. The coefficient estimate, β_{10} , in Column 1 is -7.524 (p -value < 0.05). Column 2 includes all the *focused acquisition attempts*. The coefficient estimate, β_{10} , in Column 2 is -0.338 (p -value $= 0.71$). The difference in coefficients between *diversifying* and *focused acquisition attempts* is consistent with managers in acquisition attempts being more sensitive to stock price when short selling is less impeded and when managers put more weight on market prices in deciding whether to abandon proposed acquisitions.

Table 5. Acquisition abandonment during the pre-SHO, SHO, and post-SHO periods by the type of acquisition. This table presents estimates of a triple difference analysis in a linear probability model of acquisition abandonment. The sample includes 756 value-reducing acquisition attempts by the Regulation SHO pilot and nonpilot firms announced over the period of 1 January 2002 through 31 December 2010. The dependent variable is an indicator that takes the value of one for abandoned acquisition attempts and zero for completed attempts. An acquisition attempt is classified as pilot if the acquirer was designated as a pilot firm during the Regulation SHO pilot program and is designated as SHO or post-SHO based on the announcement date of the acquisition attempt. The sample is split based on whether the acquisition attempt is categorized as a diversifying or focused acquisition attempt. A diversifying acquisition attempt is defined as one in which the target and acquiring firm are not in the same Fama–French 17 industry. Variables are defined in Appendix A. All regressions control for year and industry fixed effects. The coefficient estimates of the constant, year, and industry dummies are omitted for brevity. Heteroskedasticity-robust standard errors are clustered by industry, and t-statistics are reported in brackets. ***, **, and * indicate significance at 1%, 5%, and 10%, respectively.

	Dependent Variable: Abandonment (1,0)	
	Diversifying Acquisition Attempts	Focused Acquisition Attempts
CAR × Pilot × SHO	−7.524 ** [−2.37]	−0.338 [−0.37]
CAR × Pilot × Post-SHO	−2.310 [−1.66]	1.683 [1.63]
CAR	−1.995 * [−1.67]	−1.668 *** [−4.30]
Pilot	0.050 [1.09]	−0.019 [−0.62]
SHO	−0.035 [−0.31]	0.030 [1.24]
Post-SHO	−0.011 [−0.11]	−0.036 [−1.17]
CAR × Pilot	2.000 [1.47]	−1.178 [−1.39]
CAR × SHO	1.959 [1.62]	1.688 *** [4.03]
CAR × Post-SHO	2.232 * [1.79]	1.395 *** [3.34]
Pilot × SHO	−0.150 [−1.50]	−0.005 [−0.19]
Pilot × Post-SHO	−0.093 [−1.46]	0.060 [1.14]
Industry Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes
Observations	197	559
Adj. R ²	0.1334	0.1714

4.3.2. High-Tech vs. Non-High-Tech Target Firms

We then split the sample based on target firm characteristics. We classify observations into two samples: *high-tech target* and *non-high-tech target*. We define a *high-tech target* as one in which the target firm is identified as a high-tech firm by the SDC database and the acquirer has a different SIC Code than the target. *High-tech targets* are often young, underfunded companies without prospects for generating cash flow in the near future. For these reasons, they are more difficult to evaluate due to the uncertainty associated with values which rely on future outcomes or developments in uncharted fields (Benou and Madura (2005) and Kohers and Kohers (2000)). Thus, we assume that high-tech targets are more opaque to the acquirer and use *high-tech target* as an indicator to identify opaque targets.

We re-estimate the model in Equation (2) for each sample. Table 6 reports the results. Column 1 includes all acquisition attempts with a *high-tech target*. The coefficient estimate, β_{10} , in Column 1 is -2.334 (p -value < 0.01). Column 2 includes all the acquisition attempts with a *non-high-tech target*. The coefficient estimate, β_{10} , in Column 2 is -1.255 (p -value $= 0.27$). The difference in the coefficients between acquisition attempts with *high-tech* and *non-high-tech targets* is consistent with the proposition that managers are more sensitive to the stock price reaction when short selling is unimpeded by the uptick rule and when managers are likely to put more weight on market prices in their decisions to abandon (or complete) proposed acquisitions.

Table 6. Acquisition abandonment during the pre-SHO, SHO, and post-SHO periods by the type of target. This table presents estimates of a triple difference analysis in a linear probability model of acquisition abandonment. The sample includes 756 value-reducing acquisition attempts by the Regulation SHO pilot and nonpilot firms announced over the period of 1 January 2002 through 31 December 2010. The dependent variable is an indicator that takes the value of one for abandoned acquisition attempts and zero for completed attempts. An acquisition attempt is classified as pilot if the acquirer was designated as a pilot firm during the Regulation SHO pilot program and is designated as SHO or post-SHO based on the announcement date of the acquisition attempt. The sample is split based on whether the target firm is categorized as high-tech by the SDC database along with the acquirer having a different SIC Code than the target. Variables are defined in Appendix A. All regressions control for year and industry fixed effects. The coefficient estimates of the constant, year, and industry dummies are omitted for brevity. Heteroskedasticity-robust standard errors are clustered by industry, and t-statistics are reported in brackets. ***, **, and * indicate significance at 1%, 5%, and 10%, respectively.

	Dependent Variable: Abandonment (1,0)	
	High-Tech Target	Non-High-Tech Target
CAR × Pilot × SHO	−2.334 *** [−3.22]	−1.255 [−1.11]
CAR × Pilot × Post-SHO	−1.943 [−1.11]	0.338 [0.26]
CAR	−1.126 *** [−2.88]	−1.731 *** [−3.27]
Pilot	−0.022 [−0.25]	0.021 [0.65]
SHO	−0.062 [−0.78]	0.034 [1.53]
Post-SHO	−0.115 [−0.88]	0.000 [0.01]
CAR × Pilot	0.725 [0.82]	0.006 [0.01]
CAR × SHO	0.856 [1.23]	1.920 *** [3.88]

Table 6. Cont.

	Dependent Variable: Abandonment (1,0)	
	High-Tech Target	Non-High-Tech Target
CAR × Post-SHO	1.204 *** [3.39]	1.866 *** [2.80]
Pilot × SHO	−0.030 [−0.40]	−0.060 ** [−2.12]
Pilot × Post-SHO	−0.050 [−0.42]	0.018 [0.35]
Industry Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes
Observations	203	553
Adj. R ²	0.1048	0.1614

4.4. Corporate Governance of Acquiring Firms

In this section, we examine whether managers are more sensitive to their firm’s stock price reaction in making their decisions to abandon value-reducing acquisition attempts when short sellers are unimpeded by the uptick test and when the acquiring firm is guided by a higher level of corporate governance. Specifically, we use the *E-Index* (Bebchuk et al. 2009) as a proxy for the quality of corporate governance and separate the sample into terciles. Firms in the bottom tercile are labeled as *Strong Shareholder Rights*. Firms in the top tercile are labeled as *Weak Shareholder Rights*.

We re-estimate the model in Equation (2) for the top and the bottom terciles. Table 7 reports the results. Column 1 includes acquiring firms labeled as *Strong Shareholder Rights*. The coefficient estimate, β_{10} , is -4.387 (p -value < 0.01). Column 2 includes acquiring firms labeled as *Weak Shareholder Rights*. The coefficient estimate, β_{10} , is -1.621 (p -value < 0.05).

Table 7. Acquisition abandonment during the pre-SHO, SHO, and post-SHO periods by the corporate governance of the acquirer. This table presents estimates of a triple difference analysis in a linear probability model of acquisition abandonment. The sample includes 756 value-reducing acquisition attempts by the Regulation SHO pilot and nonpilot firms announced over the period of 1 January 2002 through 31 December 2010. The dependent variable is an indicator that is given the value of one for abandoned acquisition attempts and zero for completed attempts. An acquisition attempt is classified as pilot if the acquirer was designated as a pilot firm during the Regulation SHO pilot program and is designated as SHO or post-SHO based on the announcement date of the acquisition attempt. Column (1) includes acquisition attempts where the acquiring firm has *Strong Shareholder Rights* based on their *E-Index*, and Column (2) includes acquisition attempts where the acquiring firm has *Weak Shareholder Rights* based on their *E-Index*. Variables are defined in Appendix A. All regressions control for year and industry fixed effects. The coefficient estimates of the constant, year, and industry dummies are omitted for brevity. Heteroskedasticity-robust standard errors are clustered by industry, and t-statistics are reported in brackets. ***, **, and * indicate significance at 1%, 5%, and 10%, respectively.

	Dependent Variable: Abandonment (1,0)	
	Strong Shareholder Rights	Weak Shareholder Rights
CAR × Pilot × SHO	−4.387 *** [−4.03]	−1.621 ** [−2.26]
CAR × Pilot × Post-SHO	−0.935 [−1.55]	−0.207 [−0.49]
CAR	−2.719 *** [−7.74]	−2.007 *** [−5.95]
Pilot	0.068 [1.31]	0.065 [1.07]

Table 7. Cont.

	Dependent Variable: Abandonment (1,0)	
	Strong Shareholder Rights	Weak Shareholder Rights
SHO	−0.060 [−0.48]	0.012 [0.22]
Post-SHO	−0.151 [−1.47]	−0.053 [−1.03]
CAR × Pilot	2.486 *** [6.50]	1.704 *** [3.06]
CAR × SHO	2.758 *** [7.85]	2.167 *** [5.80]
CAR × Post-SHO	1.917 *** [3.93]	2.069 *** [5.16]
Pilot × SHO	−0.115 ** [−2.15]	−0.077 [−1.13]
Pilot × Post-SHO	0.000 [0.00]	−0.056 [−0.88]
Industry Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes
Observations	246	253
Adj. R ²	0.1883	0.1918

Though both coefficients of interest in Columns 1 and 2 are negative and statistically significant, they represent different economic significance when holding the CAR constant at its mean of −3.81%. For acquisition attempts with acquiring firms labeled as *Strong Shareholder Rights*, managers of pilot acquiring firms are 2.59 percentage points more likely to abandon the proposed acquisition than are managers of nonpilot acquiring firms. For acquisition attempts by acquiring firms labeled as *Weak Shareholder Rights*, managers of pilot acquiring firms are 1.51 percentage points more likely to abandon the proposed acquisition than managers of nonpilot acquiring firms.

These findings are consistent with the proposition that managers of pilot acquiring firms with stronger corporate governance are more sensitive to stock price changes in making the decision to abandon value-reducing acquisition attempts than are managers of nonpilot acquiring firms during the SHO period compared with the pre-SHO period. That is, managers are more sensitive to the stock price changes when short selling is less impeded by regulations and when the acquiring firm is guided by stronger corporate governance.

5. Robustness

The findings thus far are consistent with the proposition that although the participation of short sellers in stock trading often arouses the ire of managers whose firms become the targets of short selling, managers are more sensitive to stock price changes when short sellers are less impeded in playing a role in the price setting process. In this section, we address the robustness of the results using different specifications of the data and variables. All results reported are presented in the Supplementary Materials.

5.1. Delayed Response in Stock Prices Due to Impeded Short Sellers

The key premise of our study is that, as informed traders, short sellers’ unimpeded involvement in the price setting process makes the change in stock price a more accurate depiction of the value destroyed by proposed value-reducing acquisitions. Consequently, managers are more sensitive to stock price changes when short sellers are less impeded by regulation in the price setting process.

An alternative possibility is that because short sellers are impeded from participating in the price setting process, for nonpilot firms the price response is delayed. Consider an example of two acquisition attempts. Suppose that both acquirers experience a −5% CAR at the announcement of proposed acquisitions. Furthermore, suppose that the stock

of the first acquiring firm has fewer short-sale impediments than the stock of the second acquiring firm. For the second firm whose stock is more short sale impeded, the observed value reduction could be less than the actual value reduction due to lesser involvement of short sellers. In this scenario, the short sale impeded firm could experience a more gradual decline in stock price as the market converges to the actual value destruction following the announcement of the proposed acquisition. Therefore, managers of more short sale impeded firms, in expectation of further price decline following the announcement, could be more sensitive to the -5% stock price decline in the days immediately surrounding the announcement. If so, the implication is that, during the SHO period, managers of nonpilot acquiring firms would be more sensitive to stock price changes at the announcement of the proposed acquisition attempt in making their decisions to abandon (or complete) the transaction than managers of pilot firms. We find they are not, thereby ruling out this alternative possibility.

5.2. Post-Announcement Drift

Our study relies upon the validity of the three-day CAR as an unbiased measure of the value added (destroyed) due to the proposed acquisitions. One concern with this assumption is that there could be systematic drift in stock returns following acquisition announcements, in particular for firms where short selling is more impeded by regulation. If so, the difference in managers' sensitivity to stock price changes in their decisions to abandon proposed value-reducing acquisitions could be explained by the difference in stock price changes post the announcement of the proposed acquisitions. To address such a concern, we examine the difference in post-announcement drift in abnormal stock returns between pilot and nonpilot firms.

We find that for acquisition announcements made during the SHO period, there is no statistically nor economically significant difference in cumulative abnormal stock returns (CARs) between pilot and nonpilot firms in the post announcement period up to 20 trading days following the announcement of the proposed acquisitions. In particular, for acquisitions announced during the SHO period, the difference in CARs over the interval of $[+2, +20]$ for acquisitions proposed by pilot and nonpilot firms is 0.08% with a p -value of 0.46 . These results suggest that there is no difference following the announcement of acquisition attempts by pilot and nonpilot firms.

5.3. Benchmark for Value-Reducing Acquisitions

Our criterion of a CAR less than zero for the identification of value-reducing acquisition attempts implicitly assumes that CARs are measured without error. Since CARs are, indeed, measured with error, some of the attempts classified as value-reducing might be misclassified. To mitigate the concern that our results are due to misclassified value-reducing acquisition attempts, we re-estimate Equation (2) including only acquisition attempts with CARs less than -2% . Despite the decrease in sample size, the coefficient estimate of the triple interaction term $CAR \times Pilot \times SHO$ is negative and statistically significant (p -value < 0.05). This estimate indicates that the results of our analysis are not due to potentially misclassified value-reducing acquisition attempts where the CARs are slightly less than zero.

5.4. Acquirers Repeatedly Abandoning Acquisition Attempts

Some acquirers enter the sample more than once. Indeed, 67 of the sample firms abandoned value-reducing acquisition attempts three or more times. To mitigate the concern that our results are due to a handful of repeated acquirers where managers are especially sensitive to stock price in making the decision to abandon value-reducing acquisition attempts and their firms just happen to be included in the pilot program, we include only the first acquisition attempt by any acquirer and re-estimate Equation (2). Despite the decrease in sample size, the coefficient estimate of the triple interaction term $CAR \times Pilot \times SHO$ is negative and statistically significant at -2.466 (p -value < 0.10).

Thus, the results of our analysis are not due only to acquirers that repeatedly abandon value-reducing acquisition attempts.

5.5. Other Robustness Tests

We also re-estimate our models using probit and tobit specifications, using a five-day interval $(-2, +2)$ and a seven-day interval $(-3, +3)$ to calculate CAR, using the Fama–French–Carhart four factor model to calculate CAR, and including proposed acquisitions that have a transaction value of at least USD 50 million or USD 10 million. In all cases, the triple interaction term $CAR \times Pilot \times SHO$ is negative and statistically significant (all p -values < 0.05), and the triple interaction term of $CAR \times Pilot \times PostSHO$ is statistically insignificant (all p -values > 0.46).

6. Using Regulation SHO as an Exogenous Shock

Concerns about the validity of empirical tests that utilize the SHO experiment have been set forth by Litvak et al. (2020) and Heath et al. (2023). The basis of the concerns is that the causal chain argument fails as there is no evidence that the Regulation SHO pilot program significantly altered the level of short selling in pilot stocks.¹⁴ We argue that our interpretation of the results can survive such concerns.

Firstly, our study focuses on managers' perception of the manifest effect of short sellers on stock price and whether that perception of manifest short sales influences managerial actions. We presume that the Regulation SHO pilot program significantly altered managers' perception of short sellers' participation in stock price setting process. To support this presumption, we consider the number of news articles including "short sellers" or "short selling" in the *Wall Street Journal*, *New York Times*, *Washington Post*, and *USA Today* during the pre-SHO and SHO periods of the Regulation SHO pilot program. As illustrated in Figure 2, we find the monthly number of news articles including "short sellers" or "short selling" is 12.08 during the pre-SHO period and 18.13 during the SHO period. The more than 50% increase in the media attention given to short selling activities supports our presumption that the Regulation SHO pilot program exogenously increased managers' perception of the manifest effect of short sellers on stock price.

Secondly, Litvak et al. (2020) contend that the SEC erred in the setup of the regulation SHO experiment, and, thus, if researchers find a treatment effect using SHO as an exogenous shock, the effect is likely to be spurious. In particular, "the SEC took the randomly selected original nonpilot firms, picked the largest third, and suspended the uptick rule for trading in these firms' shares after regular trading hours. Moreover, the Nasdaq national market never applied short sale restrictions after trading hours". Thereby, Litvak et al. argue that the nonpilot firms in the Regulation SHO program include "partially" pilot firms.¹⁵ We acknowledge that the error in the SHO program leads us to misclassify "partially" pilot firms as nonpilot firms. In general, such a misclassification reduces the likelihood of rejecting the null hypothesis. Nevertheless, our tests do reject the null hypothesis.

Thirdly, Heath et al. (2023) comment that there were more than 40 papers that use Regulation SHO as a quasi-natural experiment, and combined, these papers examine more than 23 different dependent variables. They argue that the repeated use of a quasi-natural experiment increases the likelihood of false discoveries. They recommend that the critical values for significance tests be adjusted. Using the Heath et al. (2023) adjusted critical values, all of the coefficient estimates of $CAR \times Pilot \times SHO$ continue to be significant at the 0.05 level.

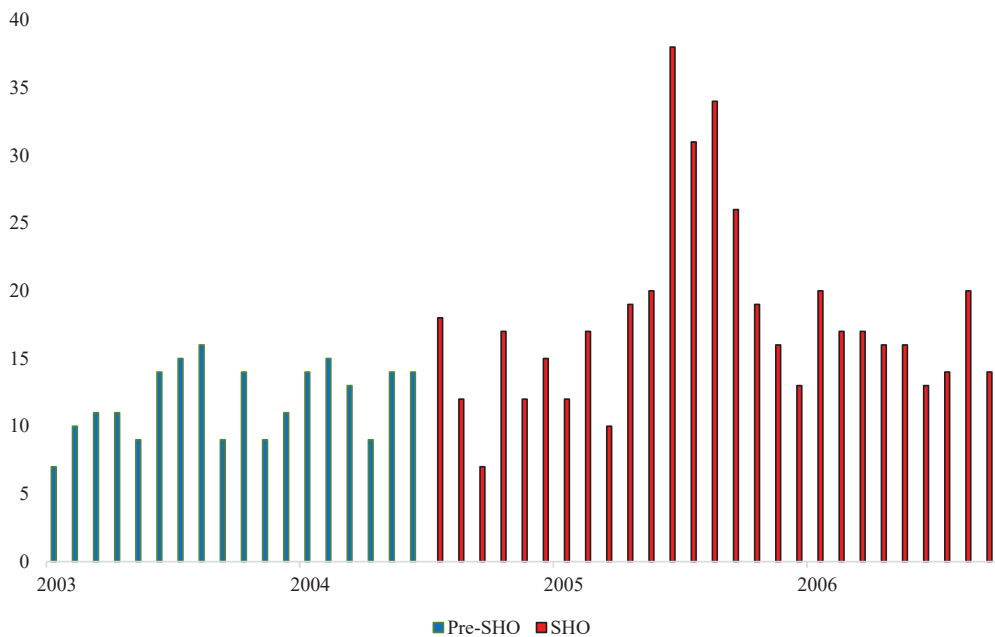


Figure 2. Media attention given to short selling activities during the pre-SHO and the SHO periods. This figure depicts the monthly number of news articles including “short sellers” or “short selling” in the *Wall Street Journal*, *New York Times*, *Washington Post*, and *USA Today* during the pre-SHO and SHO periods of the Regulation SHO pilot program. The pre-SHO period is 1 January 2002 through 1 May 2005. The SHO period is 2 May 2005 through 6 August 2007.

7. Conclusions

In this study, we exploit a quasi-randomized experiment to examine the cost of restricting short sales in the form of the reduction in market information and efficiency for managers; that is, we test whether the participation of short sellers in the price setting process increases (or decreases) managers’ sensitivity to stock price changes when making corporate decisions. Our setting is the SEC mandated exemption of short-sale uptick tests for a set of firms included in the Regulation SHO pilot program and 756 proposed value-reducing acquisition attempts during the period of 2002–2010. We find that during the pilot program, managers of pilot firms are more sensitive to stock price changes when deciding whether to abandon value-reducing acquisition attempts than nonpilot firms. Furthermore, we find that, during the pre- and post-SHO period, the sensitivity of managers to stock price changes in deciding whether to abandon (or complete) value-reducing acquisition attempts is not different between managers of pilot and nonpilot firms. We further find that managers are more sensitive to stock price changes when short sellers are less impeded in the price setting process and when managers put more weight on market prices in making corporate decisions. While this finding must include caveats based on the recent literature discussing Regulation SHO, we believe this finding is important for both corporate managers and other market participants. It is our hope that the literature in this area which examines the corporate governance role of short sellers continues to grow.

The results of our study suggest that, although the participation of short sellers in stock trading often arouse the criticism of managers whose firms become the targets of short selling, managers are more sensitive to stock price changes when short sellers are less impeded in playing a role in the price setting process. This finding can be generalized to show that despite their dislike for them, managers believe that overall market conditions and the level of information from markets is superior when short sellers are less impeded,

and therefore, the imposition of short sale constraints could have a large effect on market price information.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jrfm16110486/s1>.

Author Contributions: All authors contributed to all parts of this article. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We are grateful for the constructive and helpful comments received from Don Autore, Yingmei Cheng, Francois Cocquemas, David Denis, Diane Denis, Casey Dougal, Mara Faccio, Irena Hutton, Byoung-Hyoun Hwang, Anastasia Semykina, Henri Servaes, and Rene Stulz.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Variable Definitions

Variable	Definition
CAR	Cumulative abnormal return over the $(-1, +1)$ interval surrounding announcements of an acquisition using the market model where the announcement dates are from <i>SDC</i> .
Abandonment	One for abandoned acquisition attempts and zero for completed acquisition attempts, as reported by <i>SDC</i> .
Compete Dummy	One for acquisition attempts with the emergence of a third party who launches an offer to the same target firm while the original acquisition attempt is pending and zero otherwise, as reported by <i>SDC</i> .
Defense Dummy	One for acquisition attempts in which the target firm undertakes defensive tactics to fend off the acquisition attempt and zero otherwise, as reported by <i>SDC</i> .
Definitive Agreement Dummy	One for acquisition attempts in which the finalized terms of the deal are agreed upon prior to or on the acquisition announcement date and zero otherwise, as reported by <i>SDC</i> .
Diversifying Acquisitions	Acquisition attempts in which the target and acquiring firm are not in the same Fama–French 17 industry.
E-Index	The Entrenchment index of Bebchuk et al. (2009) in the year prior to the acquisition announcement.
Focused Acquisitions	Acquisition attempts in which the target and acquiring firm are in the same Fama–French 17 industry.
Hostile Dummy	One for acquisition attempts in which the “attitude” of the proposed acquisition is hostile and zero otherwise, as reported by <i>SDC</i> .
High-Tech Target	One for acquisition attempts in which the target firm is categorized as high-tech and the acquirer has a different SIC Code than the target and zero otherwise, as reported by <i>SDC</i> .
Lockup Dummy	One for acquisition attempts including a lockup of target shares in which the potential acquirer is granted an option to purchase shares at a fixed price even if a competing offer emerges and zero otherwise, as reported by <i>SDC</i> .
Ln(Market Cap)	The natural log of the acquirer’s market value of equity (in millions of dollars) on the 43rd trading day prior to the announcement day of the acquisition attempt.
Same-State HQ Dummy	One for acquisition attempts where the acquirer and target firms are incorporated in the same state and zero otherwise.

Variable	Definition
Tender Offer Dummy	One for acquisition attempts structured as a tender offer and zero otherwise, as reported by <i>SDC</i> .
Pre-SHO	One for acquisition attempts announced between 1 January 2002 and 1 May 2005 (inclusive) and zero otherwise.
Pilot	One for acquisition attempts where the acquirer was a pilot firm in Regulation SHO and zero otherwise.
SHO	One for acquisition attempts announced between 2 May 2005 and 6 August 2007 (inclusive) and zero otherwise.
Post-SHO	One for acquisition attempts announced between 7 August 2007 and 31 December 2010 (inclusive) and zero otherwise.
Stock Dummy	One for acquisition attempts financed or partially financed by the acquirer’s common stock and zero otherwise.
Termination Fee Dummy	One for acquisition attempts that include termination fees that the potential acquirer must pay to the target if the transaction is abandoned and zero otherwise, as reported by <i>SDC</i> .

Appendix B. Institutional Background

Short-sale price tests, introduced in the 1930s in the US equity markets, are trading restrictions that limit when a short sale can be made. One such rule, known as the “uptick test”, was adopted by the SEC in 1938 as a part of Rule 10a-1 and continued in force until 2007. According to Pessin (1978), the rule was designed to curtail “groups of speculators from pooling their capital and selling short for the sole purpose of driving down the stock price of a particular security to a level where the stockholders would panic and unload their fully owned shares”. The uptick test mandates that a short sale must be at a price either above the last traded price of the security or at the last traded price when the most recent movement between traded prices was upward. In 1994, the National Association of Securities Dealers (NASD) adopted its own stricter price test, the “bid test”, under Rule 3350. The bid test requires that a short sale occur at a price at least one penny above the bid price if the bid is a downtick from the previous bid.

On 28 July 2004, the SEC issued the Securities Exchange Act Release No. 50104 (available at <http://www.sec.gov/rules/other/34-50104.htm> (accessed on 12 November 2023)). This act suspended the operation of short-sale uptick tests for a group of stocks during a pilot period. To select the stocks in the pilot program, the SEC first sorted firms in the Russell 3000 as of June 2004 into three groups—Amex, Nasdaq nm and NYSE—excluding stocks that went public or had a spin-off between 30 April 2004 and 28 July 2004 or that were not subject to uptick tests as of 25 June 2004. The securities in each group were then ranked from highest to lowest average daily dollar volume for the period of June 2003–May 2004. In each group, every third stock was selected to be in the pilot study, yielding a total of 986 stocks that would trade without being subjected to any uptick tests during the pilot program.

As set forth by the SEC, from 2 May 2005 to 6 August 2007, pilot stocks were to be exempted from short-sale price tests. The program effectively ended one month early on 6 July 2007, when the SEC eliminated short-sale price tests for all the exchange listed stocks. The SEC stated that “[t]he general consensus was that the Commission should remove price test restrictions because they modestly reduce liquidity and do not appear necessary to prevent manipulation”.¹⁶

The decision to remove all short-sale price tests was met with hostility by managers, politicians, and the media. Just one month after the decision, Muriel Siebert, founder of the Siebert Financial Corporation and former state banking superintendent of New York, asserted that the removal of the uptick rule on short sales “may be exacerbating the downdrafts when they come along”.¹⁷ US Senator John McCain stated that the removal of the uptick rule on short sales “allowed abusive short-selling, or bearish bets on a company’s stock, to turn our markets into a casino”.¹⁸ Prominent economic commentator Jim Cramer

called for the restoration of the uptick rule and claimed that “only then will the shorts lose the upper hand, and the market will become a safer, fairer place for you to invest”.¹⁹

In April 2009, the SEC voted to seek public comment on proposals to restore a form of the uptick rule. In February 2010, the SEC adopted the alternative uptick rule by amending Rules 200(g) and 201 of Regulation SHO. The alternative uptick rule is triggered when a security’s price falls by 10% or more from the previous day’s closing price. The rule is in effect until the close of trading the next day.²⁰

Notes

- ¹ For papers on the informativeness of short sellers trades, see Miller (1977), Diamond and Verrecchia (1987), Senchack and Starks (1993), Desai et al. (2002), Duffie et al. (2002), Asquith et al. (2005), Bris et al. (2007), Boehmer et al. (2008), Saffi and Sigurdsson (2011), Engelberg et al. (2012), Boehmer and Wu (2013), Curtis and Fargher (2014), Akbas et al. (2017), and Hwang et al. (2019), among others. For papers on managers learning from prices, see Gilchrist et al. (2005), Grullon et al. (2015), He et al. (2023), among others.
- ² “Elon Musk: The Architect of Tomorrow”, Neil Strauss, 15 November 2017. Similar sentiments include, for example, Jamie Dimon’s (CEO of JP Morgan), discussion of the role of short sellers in the downfall of Bear Stearns in “S.E.C. Unveils Measures to Limit Short-Selling”, Jenny Anderson, 16 July 2008 or Patrick Byrne’s (CEO of Overstock) allegation against short sellers in “Overstock’s Phantom Menace”, Bethany McLean, 1 November 2005.
- ³ (Securities and Exchange Commission (SEC) (2007)), Economic analysis of the short sale price restrictions under the regulation SHO pilot, Office of Economic Analysis.
- ⁴ We define value-reducing acquisition attempts as those that are accompanied by a negative stock price reaction at the announcement of the deal as measured by the acquiring firm’s 3-day announcement period CAR.
- ⁵ See, for example, Luo (2005), Chen et al. (2007), Kau et al. (2008), Masulis et al. (2009), and Liu and McConnell (2013) who report that managers are more likely to abandon acquisition attempts the lower the stock returns of the acquirer at the announcement of the potential acquisition.
- ⁶ See Bebhuk et al. (2009) for further discussion on the E-Index.
- ⁷ See, for example, Boehmer and Wu (2013), Saffi and Sigurdsson (2011), Massa et al. (2015), Bris et al. (2007), Beber and Pagano (2013), and Charoenrook and Daouk (2009) among others.
- ⁸ The SEC initially published a list of 986 pilot stocks (available at <https://www.sec.gov/rules/other/34-50104.htm>) on 28 July 2004. The list was later updated to 948 pilot stocks to account for business combinations, permanent delistings, and other corporate events (available at <https://www.sec.gov/spotlight/shopilot.htm> (accessed on 13 November 2023)). For robustness, we re-estimate our models using the updated list of pilot stocks and find quantitatively similar results. These results are reported in the Supplementary Materials.
- ⁹ Missing variables include the percentage of target shares that the acquirer owns prior to the proposed acquisition, percentage of shares sought by the acquirer, transaction value, and the final transaction status (i.e., completed or withdrawn).
- ¹⁰ Following Fang et al. (2016), who also use Regulation SHO as a quasi-natural experiment, we limit our sample from 2002 (inclusive) to 2010 (inclusive). This allows us to compare the results from the three-year window when Regulation SHO was active to a three-year window prior to and post the program.
- ¹¹ See, for example, Henry (2004), Liu and McConnell (2013), Chen et al. (2015), Becht et al. (2016), and Bereskin et al. (2018) among others.
- ¹² See note 5 above.
- ¹³ We are limited in our selection of variables to proxy for the opacity of the target firm by their availability of variables for non-listed firms.
- ¹⁴ The causal chain approach sequences the results such that first order effects are tested first. The Romano and Wolf (2005) procedure is then performed resulting in updated critical values with which the second order effects are tested. This continues to the third order effects and so on until all effects have been tested. For Regulation SHO, variables related to the price formation process (i.e., manager’s perception of short sellers’ involvement) are considered second order outcomes.
- ¹⁵ “The busted randomization went unnoticed by all prior researchers, including the OEA. Indeed, Chester Spatt, the then-SEC Chief Economist who oversaw the experiment, told us that he was not aware that the SEC had created a partly treated group from the original controls”. Litvak et al. (2020).
- ¹⁶ <http://www.sec.gov/rules/other/34-50104.htm> (accessed on 12 November 2023).
- ¹⁷ See Morgenson, Gretchen. “Why the roller coaster seems wilder”, The New York Times. 26 August 2007.
- ¹⁸ See Meckler, Laura and Scannell, Kara. “McCain Says Cox Should Be Fired as SEC Chief Amid ‘Casino’ Markets” The Wall Street Journal. 18 September 2008.
- ¹⁹ See Carney, John. “Bring Back the Uptick Rule?” Business Insider. 18 November 2008.

- ²⁰ See <https://www.sec.gov/news/press/2010/2010-26.htm> (accessed on 12 November 2023) for the SEC's press release announcing the adoption of the alternative uptick rule.

References

- Aggarwal, Rajesh K., and Mufaddal. H. Baxamusa. 2013. Unrelated acquisitions. Paper presented at the American Finance Association Meeting, San Diego, CA, USA, January 3–5.
- Akbas, Ferhat, Ekkehart Boehmer, Bilal Erturk, and Sorin Sorescu. 2017. Short interest, returns, and fundamentals. *Financial Management* 46: 455–86. [CrossRef]
- Asquith, Paul, Parag A. Pathak, and Jay R. Ritter. 2005. Short interest, institutional ownership, and stock returns. *Journal of Financial Economics* 78: 243–76. [CrossRef]
- Bates, Thomas W., and Michael L. Lemmon. 2003. Breaking up is hard to do? An analysis of termination fee provisions and merger outcomes. *Journal of Financial Economics* 69: 469–504. [CrossRef]
- Bebchuk, Lucian, Alma Cohen, and Allen Ferrell. 2009. What matters in corporate governance? *The Review of Financial Studies* 22: 783–827. [CrossRef]
- Beber, Alessandro, and Marco Pagano. 2013. Short-selling bans around the world: Evidence from the 2007–09 crisis. *The Journal of Finance* 68: 343–81. [CrossRef]
- Becht, Marco, Andrea Polo, and Stefano Rossi. 2016. Does mandatory shareholder voting prevent bad acquisitions? *The Review of Financial Studies* 29: 3035–67. [CrossRef]
- Benou, Georgina, and Jeff Madura. 2005. High-tech acquisitions, firm specific characteristics and the role of investment bank advisors. *The Journal of High Technology Management Research* 16: 101–20. [CrossRef]
- Bereskin, Fred, Seong K. Byun, Micah S. Officer, and Jong-Min Oh. 2018. The effect of cultural similarity on mergers and acquisitions: Evidence from corporate social responsibility. *Journal of Financial and Quantitative Analysis* 53: 1995–2039. [CrossRef]
- Boehmer, Ekkehart, and Julie Wu. 2013. Short selling and the price discovery process. *The Review of Financial Studies* 26: 287–322. [CrossRef]
- Boehmer, Ekkehart, Charles M. Jones, and Xiaoyan Zhang. 2008. Which shorts are informed? *The Journal of Finance* 63: 491–527. [CrossRef]
- Bris, Arturo, William N. Goetzmann, and Ning Zhu. 2007. Efficiency and the bear: Short sales and markets around the world. *The Journal of Finance* 62: 1029–79. [CrossRef]
- Burch, Timothy R. 2001. Locking out rival bidders: The use of lockup options in corporate mergers. *Journal of Financial Economics* 60: 103–41.
- Chang, Eric C., Tse-Chun Lin, and Xiaorong Ma. 2019. Does short-selling threat discipline managers in mergers and acquisitions decisions? *Journal of Accounting and Economics* 68: 101223. [CrossRef]
- Charoenrook, Anchada Aida, and Hazem Daouk. 2009. *A Study of Market-Wide Short-Selling Restrictions*. Working Paper. Ithaca: Cornell University, Department of Applied Economics and Management.
- Chen, Tao, Jarrad Harford, and Chen Lin. 2015. Do analysts matter for governance? Evidence from natural experiments. *Journal of Financial Economics* 115: 383–410. [CrossRef]
- Chen, Xia, Jarrad Harford, and Kai Li. 2007. Monitoring: Which institutions matter? *Journal of Financial Economics* 86: 279–305. [CrossRef]
- Curtis, Asher, and Neil L. Fargher. 2014. Does short selling amplify price declines or align stocks with their fundamental values? *Management Science* 60: 2324–40.
- Desai, Hemang, K. Ramesh, S. Ramu Thiagarajan, and Bala V. Balachandran. 2002. An investigation of the informational role of short interest in the Nasdaq market. *The Journal of Finance* 57: 2263–87. [CrossRef]
- Diamond, Douglas W., and Robert E. Verrecchia. 1987. Constraints on short-selling and asset price adjustment to private information. *Journal of Financial Economics* 18: 277–311. [CrossRef]
- Duffie, Darrell, Nicolae Garleanu, and Lasse Heje Pedersen. 2002. Securities lending, shorting, and pricing. *Journal of Financial Economics* 66: 307–39. [CrossRef]
- Engelberg, Joseph E., Adam V. Reed, and Matthew C. Ringgenberg. 2012. How are shorts informed? Short sellers, news, and information processing. *Journal of Financial Economics* 105: 260–78. [CrossRef]
- Engelberg, Joseph E., Adam V. Reed, and Matthew C. Ringgenberg. 2018. Short-selling risk. *The Journal of Finance* 73: 755–86. [CrossRef]
- Fang, Vivian W., Allen H. Huang, and Jonathan M. Karpoff. 2016. Short selling and earnings management: A controlled experiment. *The Journal of Finance* 71: 1251–94. [CrossRef]
- Gilchrist, Simon, Charles P. Himmelberg, and Gur Huberman. 2005. Do Stock Price Bubbles Influence Corporate Investment? *Journal of Monetary Economics* 52: 805–27.
- Grullon, Gustavo, Sebastien Michenaud, and James P. Weston. 2015. The Real Effects of Short-Selling Constraints. *The Review of Financial Studies* 28: 1737–67. [CrossRef]
- He, Jie, Xiao Ren, and Xuan Tian. 2023. Do Short Sellers Affect Corporate Innovation? Evidence from a Policy Experiment. *The Review of Corporate Finance Studies*. Available online: <https://academic.oup.com/rdfs/advance-article-abstract/doi/10.1093/rdfs/cfad007/7087109?redirectedFrom=fulltext> (accessed on 12 November 2023).

- Heath, Davidson, Matthew Ringgenberg, Mehrdad Samadi, and Ingrid M. Werner. 2023. Reusing Natural Experiments. *Journal of Finance*. forthcoming. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3457525 (accessed on 12 November 2023). [CrossRef]
- Henry, Darren. 2004. Corporate governance and ownership structure of target companies and the outcome of takeovers. *Pacific-Basin Finance Journal* 12: 419–44. [CrossRef]
- Huang, Yen-Sheng, and Ralph A. Walkling. 1987. Target abnormal returns associated with acquisition announcements: Payment, acquisition form, and managerial resistance. *Journal of Financial Economics* 19: 329–49. [CrossRef]
- Hwang, Byoung-Hyoun, Baixiao Liu, and Wei Xu. 2019. Arbitrage involvement and security prices. *Management Science* 65: 2858–75. [CrossRef]
- Jennings, Robert H., and Michael A. Mazzeo. 1991. Stock price movements around acquisition announcements and management's response. *Journal of Business* 64: 139–63. [CrossRef]
- Kau, James B., James S. Linck, and Paul H. Rubin. 2008. Do managers listen to the market? *Journal of Corporate Finance* 14: 347–62. [CrossRef]
- Kohers, Ninon, and Theodor Kohers. 2000. The value creation potential of high-tech mergers. *Financial Analysts Journal* 56: 40–51. [CrossRef]
- Lamont, Owen A. 2012. Go down fighting: Short sellers vs. firms. *The Review of Asset Pricing Studies* 2: 1–30. [CrossRef]
- Litvak, Kate, Bernard S. Black, and Woongsun Yoo. 2020. The SEC's Busted Randomized Experiment: What Can and Cannot Be Learned. *Working Paper*. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2820031 (accessed on 12 November 2023).
- Liu, Baixiao, and John J. McConnell. 2013. The role of the media in corporate governance: Do the media influence managers' capital allocation decisions? *Journal of Financial Economics* 110: 1–17. [CrossRef]
- Ljungqvist, Alexander, and Wenlan Qian. 2016. How constraining are limits to arbitrage? *The Review of Financial Studies* 29: 1975–2028. [CrossRef]
- Luo, Yuanzhi. 2005. Do insiders learn from outsiders? Evidence from mergers and acquisitions. *The Journal of Finance* 60: 1951–82. [CrossRef]
- Massa, Massimo, Bohui Zhang, and Hong Zhang. 2015. The invisible hand of short selling: Does short selling discipline earnings management? *The Review of Financial Studies* 28: 1701–36. [CrossRef]
- Masulis, Ronald W., Cong Wang, and Fei Xie. 2009. Agency problems at dual-class companies. *The Journal of Finance* 64: 1697–727. [CrossRef]
- Miller, Edward M. 1977. Risk, uncertainty, and divergence of opinion. *The Journal of Finance* 32: 1151–68. [CrossRef]
- Pessin, Allan. 1978. *Fundamentals of the Securities Industry*. New York: New York Institute of Finance, p. 241.
- Romano, Joseph P., and Michael Wolf. 2005. Stepwise multiple testing as formalized data snooping. *Econometrica* 73: 1237–82. [CrossRef]
- Saffi, Pedro A., and Kari Sigurdsson. 2011. Price efficiency and short selling. *The Review of Financial Studies* 24: 821–52. [CrossRef]
- Securities and Exchange Commission (SEC). 2007. *Economic Analysis of the Short Sale Price Restrictions under the Regulation SHO Pilot*; Office of Economic Analysis. Available online: <https://www.sec.gov/news/studies/2007/regshopilot020607.pdf> (accessed on 12 November 2023).
- Senchack, A. J., Jr., and Laura T. Starks. 1993. Short sale restrictions and market reaction to short-interest announcements. *Journal of Financial and Quantitative Analysis* 28: 177–94. [CrossRef]
- Walkling, Ralph A. 1985. Predicting tender offer success: A logistic analysis. *Journal of Financial and Quantitative Analysis* 20: 461–78. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article

Knowledge Sharing and Cumulative Innovation in Business Networks

Gilles Saint-Paul

Paris School of Economics (ENS-PSL), 75014 Paris, France; gilles.saint-paul@psemail.eu

Abstract: How can we explain the success of cooperative networks of firms which share innovations, such as Silicon Valley or the Open Source community? This paper shows that if innovations are cumulative, making an invention publicly available to a network of firms may be valuable if the firm expects to benefit from future improvements made by other firms. A cooperative equilibrium where all innovations are made public is shown to exist under certain conditions. Furthermore, such an equilibrium does not rest on punishment strategies being followed after a deviation: it is optimal not to deviate regardless of another firm's actions following a deviation. A cooperative equilibrium is more likely to arise the greater the number of firms in the network. When R&D effort is endogenous, cooperative equilibria are associated with strategic complementarities between firms' research effort, which may lead to multiple equilibria.

Keywords: R&D; cooperation; innovation; growth; technical progress; information sharing; open source; Silicon Valley; cumulative knowledge

JEL Classification: O3

1. Introduction

There are several examples of the emergence of business networks or communities where firms exchange strategic information, or even share important innovations, with competitors.

One traditional example is Silicon Valley. Accounts of its success insist on the role of information sharing between firms. Information sharing allows a firm to easily improve on an innovation by another firm, thus stimulating the growth process. Employee mobility and the low enforcement of trade secrets by courts implied that many innovations were shared by firms in the network. While such a process does not seem to rely on a voluntary policy of information sharing from firms, there is evidence that the culture of Silicon Valley penalizes those entrepreneurs who do not share enough information and/or sue departing employees for violations of trade secrets. Thus, Hyde (1998) writes that “enforceability [of trade secrets] is limited because firms that litigate in defense of their trade secrets face substantial informal social and economic sanction from other firms (whose cooperation is necessary to accomplish many projects), venture capitalists, and incumbent and prospective employees”. This suggests that while the instrument of information sharing was employee mobility between firms, entrepreneurs recognized that they had an interest in cooperating with other members of the network by not blocking such mobility by litigation or other means. Another more recent example is the emergence of the “Open Source” community in the software industry. As described by its proponents (Raymond 1999), open source software allows very fast growth for a project, as improvements that different users implement cumulate. This is especially true with respect to debugging, where users who have access to the source have an interest in finding and fixing bugs, and, with enough good will, will make their fixes public.¹

This thematic is also present in the knowledge management literature, which, based on the insights of Michael Polanyi (1962), recognizes the role of inter-organizational cooperation in the production of innovation.² Similarly, Chesbrough (2003) and Chesbrough

Citation: Saint-Paul, Gilles. 2024. Knowledge Sharing and Cumulative Innovation in Business Networks. *Journal of Risk and Financial Management* 17: 137. <https://doi.org/10.3390/jrfm17040137>

Academic Editor: Thanasis Stengos

Received: 22 September 2023

Revised: 23 January 2024

Accepted: 25 January 2024

Published: 26 March 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

et al. (2014) developed the managerial concept of “open innovation” and showed that this encompasses a variety of practices that are in no way limited to software communities but are pervasive in manufacturing as well. Chesbrough’s insights has stimulated a vast empirical literature which documents the existence of business networks that openly share knowledge in a variety of sectors.³ In the 1980s, Piore and Sabel (1986) had already documented a similar knowledge-sharing and open innovation culture in the industrial districts of SMEs in northern Italy.⁴

To an economist, a simple question is: why do such networks function at all? Why do individual firms not just free-ride by using other participants’ public information while hoarding their own valuable information? To be sure, the literature on dynamic games has highlighted a number of cooperation mechanisms: “trigger strategies”, “reputation”, etc.⁵ However, these mechanisms rely on free riding being detected, which is problematic in a context where cooperation is about information revelation, since one may well play cooperative while having nothing valuable to reveal.

In this paper, I describe another, far more robust, cooperation mechanism which applies to the revelation of innovations and captures part of the incentives underlying the open source movement. I assume innovations are cumulative, i.e., an innovation opens the door for further innovations. An individual firm then has an incentive to make its innovations public, because this will increase the number of firms heading toward the next step in the technological ladder. If it is expected that the next innovation will also be made public, sharing one’s invention with other members in the network will shorten the time until the next innovation arrives, which benefits the individual firm. As is shown below, this benefit may outweigh the short-run cost of lower profits for the innovating firm.

This mechanism does not rely on cheating being detected nor on the sustainability of a punishment strategy. By not revealing its information, the firm punishes itself, since it will have to wait longer for the next innovative step. In other words, the losses from non-cooperation are genuine and do not rely on the other players’ strategic responses to one’s deviation. However, given that the benefits from information sharing rely on future innovations also being public, the argument does rest on the horizon being infinite and a “Nash” equilibrium where information is kept private always exists. Nevertheless, for a range of parameters, a cooperative equilibrium also exists. Another interesting result is that the cooperative equilibrium is more likely to exist the larger the number of firms in the network. This contrasts with traditional arguments, which suggest that free-riding is more likely the greater the number of agents. Here, more firms mean a much greater benefit of making one’s innovation public, because a much larger number of firms will work toward the next step.

I also show that the cooperative equilibrium has faster growth than the non-cooperative equilibrium. In some sense, when innovations are cumulative, the case for public innovations is stronger than when they are not, when it can be shown that public innovations may reduce long-run growth on net because they depress the profitability of private innovation (Saint-Paul 2003).

Finally, the model is extended to allow for an endogenous research and development (R&D) effort. It is shown that the results are robust to introducing such endogeneity. Furthermore, as the cumulative benefits of an innovation depend on improvements by other firms, and thus on their R&D effort, research efforts by different firms are now complementary, which may lead to multiple equilibria.

Since the seminal work of Schumpeter (1942), the endogenous growth literature has paid a lot of attention to the effect of competition on innovation and growth (see Romer (1990); Aghion and Howitt (1992); Grossman and Helpman (1993)). A subset of that literature has analyzed the value of R&D cooperation between firms when there are spillovers. The value of technology sharing is well recognized by that literature. Hence, Baumol (1992) writes that “in an industry with, say, ten firms similar in output and investment in R&D, each member of a nine-firm technology cartel can expect to obtain immediate access to nine times the number of innovations that the remaining enterprise can anticipate on the

average". However, to my knowledge, this is the first paper to analyze how that advantage can be turned into a cooperation mechanism when innovation is cumulative. The literature, instead, typically relies on traditional trigger strategy mechanisms, exogenously assumes cooperation, or uses a cooperative game framework.⁶ Furthermore, most of it deals with cooperation in setting R&D levels rather than in the decision to make an innovation public.⁷ For example, Cozzi (1999) analyzes the consequences for growth of switches between cooperative and non-cooperative equilibria, but cooperation is enforced by a standard trigger-strategy mechanism, and cooperation is about the R&D level. Another strand of the literature (d'Aspremont and Jacquemin 1988; Leahy and Neary 1997) compares outcomes with and without R&D cooperation when there are spillovers, but again, it does not ask the question of how cooperation is enforced and focuses on investment rather than information. Finally, a series of works by Scotchmer (1991, 1996, 2004) discusses conflicts over intellectual property rights when innovation is cumulative and the efficiency properties of different patent systems in this case.

The paper which is most related to the current one is Bessen and Maskin (2008). In that paper, as in here, a given innovation exerts positive spillovers on further innovations. Bessen and Maskin perform a welfare analysis of intellectual property regimes and show that under sequential innovation, no patents may be preferable to IP protection. But, in their paper, if IP exists, it is always privately optimal to use it. In the present paper, it is actually an equilibrium outcome for firms to voluntarily disclose their innovation regardless of the IP regime.

2. The Model

2.1. The Intratemporal Profit Function

We consider a specific sector of an economy, which is homogeneous enough and where firms are large enough relative to the total size of the sector, for the economic effects analyzed here to be relevant. Such a sector could be, say, Silicon Valley, the agroforestry sector in Brazil, or flexible manufacturing garment firms in the Bologna area. In that sector, there are N firms, competing with each other by producing differentiated goods. Each firm i is characterized by a level of technological advancement captured by an integer number n_i . The profit of firm i is given by a function

$$\pi_i(n_1, \dots, n_i, \dots, n_N).$$

We assume that technical progress in a given firm increases its profits:

$$\frac{\partial \pi_i}{\partial n_i} > 0,$$

while technical progress in another firm decreases profits:

$$\frac{\partial \pi_i}{\partial n_j} < 0, j \neq i. \quad (1)$$

Finally, the progression of all firms by one step in the technological ladder increases the profits of any given firm:

$$\sum_{j=1}^N \frac{\partial \pi_i}{\partial n_j} > 0.$$

We shall assume that if all firms other than i have the same technical level n , while firm i has technical level $\hat{n}$, then firm i 's profit can be expressed in the following way:

$$\begin{aligned} \pi_i(n, \dots, n, \hat{n}, n, \dots, n) &= \pi(\hat{n}, n) \\ &= A^n f(\hat{n} - n), \end{aligned} \quad (2)$$

where by assumption, $A > 1$, $f(\cdot) > 0$, and

$$f(q+1)/f(q) > A. \quad (3)$$

The latter inequality is needed for (1) to hold.

Equation (2) expresses the firm's profit as an increasing function of the other firm's overall technological level, as captured by n , and the gap between firm i and the other firms, as captured by $\hat{n} - n$. For competitive effects to dominate, firm i has to suffer when other firms innovate, which will be true if the second effect is stronger than the first effect, as captured by our assumption that $f(q+1)/f(q) > A$. In economic terms, this condition sets a lower bound on the value of climbing the technological ladder. In Proposition 1 below, an upper bound is assumed and shown to be a sufficient condition for an "open innovation" equilibrium to exist.

Example 1. The demand curve for firm i is isoelastic and given by $y_i = Yp^{-\alpha}(p_i/p)^{-\sigma}$, where $\sigma > 1$, Y is an index of demand for the whole sector and p is a sectoral price index given by

$$p = \left(\sum_{i=1}^N p_i^{1-\sigma} \right)^{\frac{1}{1-\sigma}}.$$

Firm i 's unit cost is c/B^{n_i} , with $B > 1$; thus, each step in technical progress represents a geometric reduction in unit costs. Firms set their price so as to maximize their profit $\pi_i = y_i(p_i - c/B^{n_i})$. They neglect the impact of their decision on the sectoral price level and thus set $p_i = (\sigma - 1)/\sigma \cdot c/B^{n_i}$. Thus, we find that the equilibrium $\pi_i(\dots)$ function is given by

$$\pi_i(n_1, \dots, n_N) = KB^{(\sigma-1)n_i} \left[\sum_{i=1}^N B^{(\sigma-1)n_i} \right]^{\frac{\sigma-\alpha}{1-\sigma}},$$

where K is a constant given by $K = Y \left[\frac{\sigma c}{\sigma-1} \right]^{\frac{\sigma-\alpha}{1-\sigma}} \frac{c^{1-\sigma}}{\sigma-1} \left(\frac{\sigma}{\sigma-1} \right)^{-\sigma}$. If $n_j = n$, $j \neq i$, and $n_i = \hat{n}$, we obtain

$$\pi_i = KB^{n(\alpha-1)} \left[B^{(\sigma-1)(\hat{n}-n)} (N-1 + B^{(\sigma-1)(\hat{n}-n)})^{\frac{\sigma-\alpha}{1-\sigma}} \right];$$

thus we have

$$A = B^{\alpha-1}$$

and

$$f(q) = K \left[B^{(\sigma-1)q} (N-1 + B^{(\sigma-1)q})^{\frac{\sigma-\alpha}{1-\sigma}} \right].$$

The properties $A > 1$ and $\partial \pi_i / \partial n_j < 0$, $j \neq i$ hold if $1 < \alpha < \sigma$. One can then check that the property $f(q+1)/f(q) > A$ then holds, since

$$\frac{f'(q)}{f(q)} - \ln A = (\sigma - \alpha) \ln B - \frac{(\sigma - \alpha) B^{(\sigma-1)q} \ln B}{(N-1 + B^{(\sigma-1)q})} > 0,$$

which is positive if $\sigma > \alpha$.

This example shows how a functional form like (2) can be derived from monopolistic competition among differentiated goods. As the whole sector climbs the technological ladder (i.e., the n_i values grow), it grows in terms of output and, as long as $\alpha > 1$, in terms of profits.

Without loss of generality, we shall normalize $f(0)$ to 1 (in the previous example, this amounts to picking the right value of Y , so that $K = N^{\frac{\sigma-\alpha}{\sigma-1}}$).

Remark 1. Going back to (2), note that A is the proportional effect on a firm's profit of an increase in the technological level of all firms by one step provided all other firms have the same technical level. Conversely, the effect on a firm of a unit increase of its technical level, holding that of other firms constant, is $f(\hat{n} + 1 - n) / f(\hat{n} - n)$, which is greater than A . Growth in a single firm must increase that firms' profits by more than growth in all firms; otherwise, it would not be harmed by growth in other firms and there would not be any coordination problem. However, in the example we just constructed, we have

$$\lim_{q \rightarrow +\infty} \frac{f(q+1)}{f(q)} = B^{\alpha-1} = A. \quad (4)$$

In other words, this mathematical property means that asymptotically, if a firm become infinitely productive relative to the others, thus covering the whole market, each innovation will increase its profits by the same factor A , which increases a firm's profits when an innovation is implemented by all firms. This property is quite natural: if a firm becomes infinitely productive relative to others, it is as if one has a single firm in the market. The property should therefore hold in any model where, asymptotically, the proportional effect of an innovation which is adopted by all firms on profits does not depend on the number of firms. In other words, property (4) is not a knife-edge case but should be expected to generally hold. As shown below, property (4) plays an instrumental role in making the upper bound on individual innovation gains, that delivers the cooperative outcome in Proposition 1, compatible with Assumption (3).

2.2. The Dynamics of Innovation

The preceding subsection has introduced the intratemporal profit function. We now describe the dynamics of innovation.

Time is continuous; firms maximize the expected present discounted value of their profits given by

$$V_i(t) = E_t \int_t^{+\infty} \pi_i(n_1(u), \dots, n_i(u), \dots, n_N(u)) e^{-ru} du, \quad (5)$$

where r is the discount rate, assumed fixed and exogenous. In order for the value of a firm to be well defined, we shall assume that r is larger than a firm's maximum potential growth rate. That is,

$$\lambda N(A - 1) < r \quad (6)$$

The technical level available to a given firm jumps from n_i to $n_i + 1$ according to a Poisson process with arrival rate λ . At this stage of analysis, this process is assumed exogenous, and, therefore, so is parameter λ . Despite that, growth remains endogenous because the technical level of a given firm is transferable to other firms. Firm i has the option of revealing its level to other firms, in which case they can jump to level n_i , i.e., in some sense, adopt its technology. Therefore, with probability λ per unit of time, a given firm's technology level improves by one step, which allows it to increase it from n_i to $n_i + 1$, and it has two options:

1. Do not share its innovation with other firms, in which case they all remain at level n_j .
2. Share the innovation, in which case all firms such that $n_j < n_i + 1$ can upgrade to level $n_i + 1$.

To fix ideas, we shall assume that an innovation can be released only at the time it occurs and not after, although this is immaterial and only simplifies the analytics.

Thus, the only decision to be made by firms in this setting is whether or not to make an innovation public at the time it arrives. We will relax it in Section 4, where the R&D effort and the arrival rate of innovations are made endogenous.

2.3. The Nash Solution

Clearly, there always exists an equilibrium where firms do not share innovations. If I anticipate that other firms will never share their innovations with me, giving away my innovations to them will increase their technological level forever, which reduces my own profits, by virtue of the fact that $\partial \pi_i / \partial n_j < 0$. Thus, it is a subgame perfect equilibrium to never release information. In such an equilibrium, the average speed of innovation for any given firm is λ . As in the literature on repeated games (see Fudenberg and Tirole 1991), this subgame perfect equilibrium is the repeated Nash equilibrium of the static game where players ignore the future consequences of current outcomes.

3. Cooperative Equilibria

We now construct an equilibrium such that it is profitable for firms to reveal their innovations to their competitors, despite the fact that it reduces their profits upon impact. The equilibrium concept is again the standard subgame perfect equilibrium.⁸

We construct an equilibrium where an innovation is shared every time it occurs. Thus, in the equilibrium path, all firms have the same technological level. Situations with a dispersion in technological levels only occur off the equilibrium path.

To construct such an equilibrium, we compute the value of a firm along the equilibrium path and an upper bound of the value of deviating by not releasing information; we then show that the former is greater than the latter for a given range of parameters. This yields the following proposition:

Proposition 1. Assume $f(\cdot)$ satisfies

$$\sup_{k=0,1,\dots} A^{-k} f(k) \leq \frac{r + \lambda N}{r + \lambda(A - 1) + \lambda N(2 - A)} \quad (7)$$

Then, there exists a equilibrium where each innovation is instantaneously made public.

Proof. Along the equilibrium path which we seek to construct, all firms have the same technological level n . Furthermore, all firms upgrade to level $n + 1$ with a flow probability equal to λN , since the day any given firm innovates, its innovation spreads to the whole sector. Consequently, the value of a firm when the whole sector is at technical level n is determined by

$$V_C(n) = \frac{\pi(n, n) + \lambda N V_C(n + 1)}{r + \lambda N}.$$

The numerator is equal to the sum of the profit flow at technical level n , $\pi(n, n)$, and of the flow probability of any one firm innovating, equal to the product of the arrival rate for innovations λ and the number of firms N , times the value of the firm when all firms upgrade to the next level, $V_C(n + 1)$. Iterating forward allows for computing the value of a firm along the equilibrium path:

$$V_C(n) = \sum_{i=0}^{+\infty} \left(\frac{\lambda N}{r + \lambda N} \right)^i \frac{\pi(n + i, n + i)}{r + \lambda N},$$

which, using (2), is equivalent to

$$V_C(n) = \frac{A^n}{r - \lambda N(A - 1)} \quad (8)$$

For this formula to be meaningful, it must be that $\lambda N(A - 1) < r$, i.e., that the growth rate of profits be lower than the interest rate; otherwise, the value of a firm would be infinite.

Let $\bar{f} = \sup_{k=0,1,\dots} A^{-k} f(k)$. Then:

$$\begin{aligned} \pi_i(n_1, \dots, n_i, \dots, n_N) &\leq \pi_i(0, \dots, n_i, \dots, 0) \\ &= f(n_i) \\ &\leq A^{n_i} \bar{f}. \end{aligned} \quad (9)$$

Let $\vec{n} = (n_1, \dots, n_N)$ denote a vector of firm's technological advancements, and $V_i(\vec{n})$ denote the maximum value of firm i , over all possible courses of action, if $\vec{n}$ is the current state of technology level⁹.

Let

$$\bar{V}_i(n, p) = \max_{\substack{\vec{n}=(n_1, \dots, n_N) \\ n_i=n \\ n_j \leq p, j \neq i}} V_i(\vec{n})$$

Note that We define operator T_j as follows. If $\vec{n} = (n_1, \dots, n_N)$, then

$$T_j \vec{n} = (\max(n_1, n_j), \dots, \max(n_k, n_j) \dots, \max(n_N, n_j)).$$

Operator T_j tells us how the state vector is transformed if firm j releases its technical level: all firms with a lower technical level are lifted to level n_j . We also define U_j , which tells us how the state vector is transformed upon a shock hitting firm j , before it has decided whether to make it public or not:

$$U_j \vec{n} = (n_1, \dots, n_{j-1}, n_j + 1, n_j, \dots, n_N).$$

The value function $V_i(\vec{n})$ then obeys the following recursive condition:

$$V_i(\vec{n}) = \frac{\pi_i(\vec{n}) + \lambda \sum_{j=1}^N \max(V_i(T_j U_j \vec{n}), V_i(U_j \vec{n}))}{r + \lambda N}.$$

For any $\vec{n} = (n_1, \dots, n_N)$ such that $n_j \leq p < n_i = n$, $j \neq i$, we have, for $j \neq i$, $V_i(T_j U_j \vec{n}) \leq \bar{V}_i(n, p+1)$ and $V_i(U_j \vec{n}) \leq \bar{V}_i(n, p+1)$, while $V_i(U_i(\vec{n})) \leq \bar{V}_i(n+1, p)$ and $V_i(T_i U_i(\vec{n})) \leq \bar{V}_i(n+1, n+1)$. This, along with (9), implies

$$V_i(\vec{n}) \leq \frac{A^{n_i} \bar{f} + \lambda(N-1) \bar{V}_i(n, p+1) + \lambda \bar{V}_i(n+1, n+1)}{r + \lambda N},$$

consequently:

$$\forall p < n, \bar{V}_i(n, p) \leq \frac{A^n \bar{f} + \lambda(N-1) \bar{V}_i(n, p+1) + \lambda \bar{V}_i(n+1, n+1)}{r + \lambda N}. \quad (10)$$

To proceed, we now show that for $\vec{n} = (n_1, \dots, n_N)$ such that $n_j \leq p \leq n_i = n$, it must be that

$$V_i(\vec{n}) \leq \frac{A^n \bar{f}}{r - \lambda N(A-1)}.$$

To see this, write

$$V_i(\vec{n}) = \max_{\vec{n}(\cdot) \in \Phi} \int_0^{+\infty} \int_{\Omega_t} \pi_i(\vec{n}(\omega)) dF(\omega, t) e^{-rt} dt,$$

where ω is a state of nature representing the whole history of shocks between date 0 and date t , $\vec{n}(\omega)$ is the corresponding state vector (which depends on the strategies followed by firms), and $dF(\omega, t)$ is the probability distribution of ω at date t . Φ is the set of feasible strategies, which is represented by a mapping defining the current state vector as a function of the whole history of shocks. Clearly, regardless of which strategy is followed, the most advanced firm climbs by at most one step every time a firm is hit by a shock. Thus,

if $\vec{n}(\omega) = (n_1(\omega), \dots, n_N(\omega))$, then $\max_k n_k(\omega) \leq n + S(\omega)$, where $S(\omega)$ is the total number of shocks that has occurred to any firm between 0 and t . In other words, $\Phi \subset \{\vec{n}(\cdot), \max_k n_k(\omega) \leq n + S(\omega)\}$. $S(\omega)$ follows a Poisson distribution with arrival rate λN (that is, $P_t(S(\omega) = k) = \frac{(\lambda N t)^k}{k!} e^{-\lambda N t}$). Consequently, we have in particular that $n_i(\omega) \leq n + S(\omega)$ so that $\pi_i(\vec{n}(\omega)) \leq A^{n+S(\omega)} \bar{f}$. Hence:

$$\begin{aligned} V_i(\vec{n}) &\leq A^n \bar{f} \int_0^{+\infty} \int_{\Omega_t} A^{S(\omega)} dF(\omega, t) e^{-rt} dt \\ &= A^n \bar{f} \int_0^{+\infty} \left(\sum_{k=0}^{+\infty} A^k \frac{(\lambda N t)^k}{k!} e^{-\lambda N t} \right) e^{-rt} dt \\ &= A^n \bar{f} \int_0^{+\infty} e^{-(r - \lambda N(A-1))t} dt \\ &= \frac{A^n \bar{f}}{r - \lambda N(A-1)}. \end{aligned}$$

Note that this inequality holds regardless of the strategies that are followed by firm i or any other firms with respect to revealing their technologies upon being hit by an innovation. A corollary is that

$$\bar{V}_i(n, p) \leq \frac{A^n \bar{f}}{r - \lambda N(A-1)} \text{ for } p \leq n. \quad (11)$$

Substituting that inequality into the second term in the numerator of the right-hand side (RHS) of (10) (which we can do since in (10) one assumes $p < n$), we obtain:

$$\bar{V}_i(n, p) \leq \frac{A^n \bar{f} + \lambda(N-1) \frac{A^n \bar{f}}{r - \lambda N(A-1)} + \lambda \max \frac{A^{n+1} \bar{f}}{r - \lambda N(A-1)}}{r + \lambda N}. \quad (12)$$

Rearranging, we get that

$$\bar{V}_i(n, p) \leq \frac{A^n \bar{f}}{r - \lambda N(A-1)} \frac{r - \lambda N A + 2\lambda N - \lambda + A\lambda}{(r + \lambda N)}.$$

Now, if inequality (7) holds, then $\bar{V}_i(n, p) \leq V_i(n, \dots, n) = V_C(n) = \frac{A^n}{r - \lambda N(A-1)}$. Consequently, releasing innovation always dominates hoarding it when other firms are expected to follow the equilibrium path, which shows that it is indeed an equilibrium. $\square$

An important aspect of proposition 1 is that a cooperative equilibrium is more likely to exist the greater the number of firms participating in the network (holding the $f(\cdot)$ function invariant). This effect runs counter to the usual analysis of free rider problems, where cooperation is made more difficult by a greater number of players. Here, the greater the number of players, the quicker one will be paid back for sharing one's innovations with others, as the next technological step is discovered by one participant.

Since $A > 1$, it is also clear that the RHS of (7) is an increasing function of λ : the greater the frequency of innovations, the more likely it is that a cooperative equilibrium exists. Clearly, a greater arrival rate of innovations brings ahead in time the benefits of further innovations from other firms. Perhaps this explains why open innovation plays an important role in infant industries with a steep learning curve such as electric vehicles (see Purificato 2014; Cabigiosu 2023).

Note also that the proof relies on using an upper bound for the value of deviating for any response of other players to the deviation. Consequently, cooperation is not sustained by a trigger punishment strategy, as deviation is deterred even in the case where all other firms continue to release their innovations at all nodes following the deviation. The loss from other firms being one step backwards relative to cooperation is enough to deter opportunistic behavior. In some sense, the feedback effects of benefitting from other's

future improvements on one’s innovation generates gains from cooperation that are more robust—in that a firm contemplating deviations can ignore what others would do following its deviation—than those coming from punishment strategies. Thus, even if for some reason a deviation was undetected¹⁰, incentives to cooperate would still remain.

Note that for inequality (7) to hold and be compatible with the restriction that $f(q+1)/f(q) > A^{11}$, it must be that $\lim_{q \rightarrow +\infty} f(q+1)/(f(q)A) = 1$. Otherwise, the left-hand side (LHS) of (7) could not be bounded and (7) would not be satisfied. However, as argued above, $\lim_{q \rightarrow +\infty} f(q+1)/(f(q)A) = 1$ is a rather plausible condition¹²; if $\lim_{q \rightarrow +\infty} f(q+1)/(f(q)A) > 1$, then a single firm would grow faster than N firms for the same arrival rate of innovation, and we are not able to put a bound on the value of deviating from cooperation, although such a bound, and a cooperative equilibrium, may still exist.

Going back to the isoelastic example discussed above, in that special case, we have

$$\begin{aligned} \sup A^{-k} f(k) &= \lim_{k \rightarrow +\infty} A^{-k} f(k) \\ &= \lim_{k \rightarrow +\infty} KB^{(\sigma-\alpha)q}(N-1+B^{(\sigma-1)q})^{\frac{\sigma-\alpha}{1-\sigma}} \\ &= K \\ &= N^{\frac{\sigma-\alpha}{\sigma-1}}, \end{aligned}$$

where the last equality comes from the normalization $f(0) = 1$.

Thus, a sufficient condition for a cooperative equilibrium to exist is

$$N^{\frac{\sigma-\alpha}{\sigma-1}} < \frac{r + \lambda N}{r + \lambda(A-1) + \lambda N(2-A)}. \tag{13}$$

A first point to be made is that the RHS grows with N and is equal to 1 for $N = 1$. Therefore, for any given N , condition (13) is always satisfied for $(\sigma - \alpha)/(\sigma - 1)$ to be low enough. At face value, it seems that when N becomes large, the inequality is violated. This suggests that taking into account the effect of N on $f()$ may reverse the result that large values of N favor cooperation. This is due to an additional effect: with a larger number of differentiated firms, having an edge on the others is more valuable for a single firm, as the effect of decreasing marginal revenues at the sector level (captured by parameter α) is less severe. However, too large values of N are not possible, because (6) must hold. For intermediate values of N , it is difficult to show whether or not (13) holds analytically, but numerical computations suggest that it is more likely to hold for large, acceptable values of N than for small values of N . A typical example is given in Table 1. Therefore, the conclusion that larger networks make cooperation easier remains true despite the countervailing additional effects that were ignored when $f()$ was treated as exogenous.

Table 1. Cooperation range, $A = 1.8$, $\lambda = 0.015$, $\sigma = 2$, $\alpha = 1.8$.

r	Maximum Value of N	Range over Which (13) Holds
0.04	3	2–3
0.05	4	2–4
0.06	5	2–5
0.07	5	2–5
0.08	6	4–6
0.1	8	6–8
0.13	10	9–10
0.16	13	13–13
0.18	15	–

We can also establish a necessary condition for cooperation to hold, which has similar determinants as the sufficient one derived in Proposition 1.

Proposition 2. A necessary condition for a cooperative equilibrium to exist is

$$f(1)/A \leq \frac{r - \lambda(A - 1)}{r - \lambda N(A - 1)}. \tag{14}$$

Proof. Assume a cooperative equilibrium exists and consider a firm which deviates by not releasing its innovation. Assume it releases its innovation the next time it has one. Then, this strategy yields a value given by

$$\tilde{V}(n + 1, n) = \frac{A^n f(1) + \lambda(N - 1) \frac{A^{n+1}}{r - \lambda N(A - 1)} + \lambda \frac{A^{n+2}}{r - \lambda N(A - 1)}}{r + \lambda N}.$$

The middle term in the numerator reflects the fact that if another firm is hit by an innovation, it will release it, and all firms will be at technical level $n + 1$.

For a cooperative equilibrium to exist, this strategy must yield a lower value than releasing the innovation, i.e., one must have

$$\tilde{V}(n + 1, n) \leq \frac{A^{n+1}}{r - \lambda N(A - 1)}.$$

One can straightforwardly check that it is equivalent to (14). $\square$

Condition (14) becomes satisfied very easily as N grows. Therefore, it rules out cooperative equilibria only for fairly low values of N . In the case of our isoelastic example, for the parameter values of Table 1, it always holds. The following table (Table 2) shows the range of values of N where it does not hold, for a different set of parameters, that differ from Table 1 only in that α is lower. It confirms the general message that cooperation is more likely with more firms.

Table 2. Non-cooperation range, $A = 1.8$, $\lambda = 0.015$, $\sigma = 2$, $\alpha = 1.3$.

r	Maximum Value of N	Range over Which (14) Is Violated
0.04	3	—
0.05	4	2–2
0.06	5	2–2
0.07	5	2–3
0.09	7	2–4
0.1	8	2–5
0.12	10	2–6
0.14	11	2–8

4. Endogenous R&D Effort: The Role of Strategic Complementarities

In the preceding analysis, the only decision made by a firm is whether or not to release its innovation. The arrival rate of innovations λ is entirely exogenous. It is relatively straightforward to extend the model to allow for an endogenous λ and to compute its equilibrium value in a cooperative equilibrium. The cost is that the sufficient condition that I am able to establish for such an equilibrium to exist is more stringent than in the previous section.

Endogenizing λ yields two key insights. First, one may believe that the previous result that an increase in the network’s size boosts growth and is good for sustaining cooperation can be overturned with an endogenous λ . As N is greater and innovations arrive at a higher rate, an individual firm might be tempted to spend less on R&D and reduce λ . In fact, that intuition is incorrect: at a cooperative equilibrium¹³, an increase in N increases λ locally. The reason is that the total arrival rate of innovations is additive in each firm’s specific λ , so that an increase in the number of firms does not reduce the marginal gain

from increasing one's λ . On the contrary, a greater value of N increases the incentives for R&D via a capitalization effect. The greater the N value, the faster the rate at which the increments in profits from a given innovation grow, and the greater the incentives to innovate. Or, to put it otherwise, the greater the N value, the greater the speed at which my innovation is improved by other firms, and the greater my incentive to innovate.

Second, there exist strategic complementarities between the R and D effort of different firms. The increase in R&D by one firm tends to increase R&D by another firm. The reason is again the capitalization effect. If other firms increase their R&D effort, improvements on my innovation will come faster, and my incentive to innovate is larger. As a result, there may in principle be several cooperative equilibria with different arrival rates of innovation.

To endogenize the arrival rate of innovations, I assume that at each point in time a firm i can choose its own value of λ at a cost equal to $\pi_i c(\lambda)$, where π_i is its current profit. That the cost is proportional to π_i ensures that it will not become negligible relative to benefits as the economy grows. As cooperative equilibria are symmetrical, it is easy to compute the equilibrium common value of λ in such an equilibrium. Denoting again the value of a firm when all have a technical level equal to n by $V_C(n)$, the Bellman equation for an individual firm can be written as

$$V_C(n) = \max_{\lambda} \frac{\pi(n, n)(1 - c(\lambda)) + \hat{\lambda}(N - 1)V_C(n + 1) + \lambda V_C(n + 1)}{r + \lambda + \hat{\lambda}(N - 1)},$$

where $\hat{\lambda}$ is the common value of λ of other firms, which is taken as given by the firm. We assume that $c(\lambda)$ is concave, differentiable, and increasing over $[\underline{\lambda}, \bar{\lambda}]$, with $c'(\underline{\lambda}) = 0$, $c(\bar{\lambda}) = 1$. If there is an interior solution for λ , it is given by the first-order condition:

$$c'(\lambda)\pi(n, n) = V_C(n + 1) - V_C(n). \quad (15)$$

This equation has the usual straightforward interpretation. The LHS is the marginal cost of increasing λ by one unit. The RHS is the marginal benefit, which is equal to the capital gain made when all firms climb one step in the technological ladder. In a symmetrical equilibrium, we have $\lambda = \hat{\lambda}$, and the equivalent of Equation (8) holds, i.e.,

$$V_C(n) = \frac{A^n(1 - c(\lambda))}{r - \lambda N(A - 1)}. \quad (16)$$

Substituting that along with (2) into (15), and normalizing again $f(0)$ to 1, we obtain an equation determining the equilibrium value of λ , which is denoted by λ^* :

$$c'(\lambda^*) = \frac{(A - 1)(1 - c(\lambda^*))}{r - \lambda^* N(A - 1)}. \quad (17)$$

If $(\frac{r}{N(A-1)}) > 1$, then this equation always has at least one solution, which is indeed interior and satisfies (15).¹⁴

While the LHS is a increasing function of λ , the RHS may be either increasing or decreasing. This is because a rise in λ has two conflicting effects on the capital gains from an innovation. Because R and D costs are proportional to profits, a higher λ compresses the difference in income flows between two consecutive technological levels: one can call that a revenue effect. On the other hand, a higher λ increases the growth rate of that difference, which tends to increase its expected present discounted value: this is the capitalization effect. If the capitalization effect is strong enough, then multiple equilibria may arise, as illustrated in Figure 1.

In such a case, as in the general analysis of Cooper and John (1988), multiple equilibria come from a strategic complementarity between a firm's research effort and the effort of other firms, as discussed above. Following Cooper and John, we can check that equilibria are Pareto rankable from the point of view of the firms¹⁵. Consider two equilibria, denoted

by subscripts 0 and 1, such that $\lambda_1^* > \lambda_0^*$. Using the value function (16), we see that Equilibrium 1 dominates if

$$\frac{(1 - c(\lambda_1^*))}{r - \lambda_1^* N(A - 1)} > \frac{(1 - c(\lambda_0^*))}{r - \lambda_0^* N(A - 1)}.$$

Substituting the first-order condition (17), we see that this is equivalent to $c'(\lambda_1^*) > c'(\lambda_0^*)$, which clearly holds. Equilibria with higher R&D effort clearly dominate.

The preceding discussion does not tell us, however, whether a cooperative equilibrium exists. But, as shown in Appendix A, Proposition 1 can be extended and a non-empty sufficient condition for a cooperative equilibrium can be established, although this condition is much less tractable than the one derived in Proposition 1.

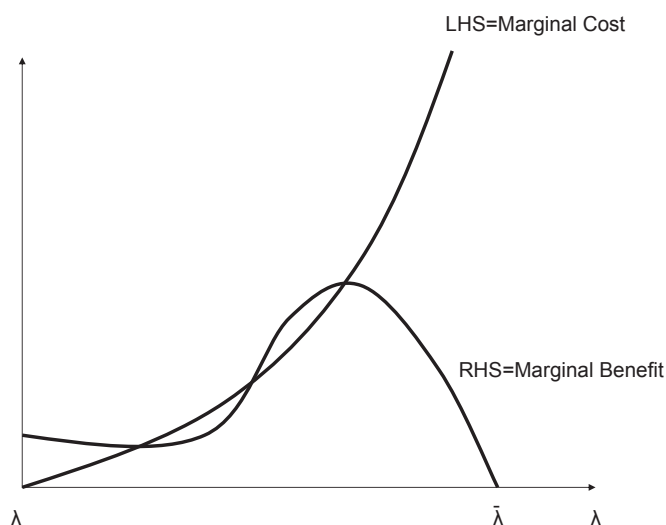


Figure 1. Multiple equilibrium R&D effort.

5. Conclusions

We hope the present paper has shed some light on the viability of public innovation when technical progress has a cumulative dimension. We have shown the existence of a “cooperative” equilibrium in which it is a subgame perfect strategy for firms in a given industry to share their innovations with others, because they expect to benefit from the incremental innovations that other firms will make, building on its own knowledge. Meanwhile, as in the literature on cooperation in dynamic games, there also exists a non-cooperative equilibrium in which innovation is not shared; here, the cooperative equilibrium is somewhat more robust than generally in that literature. Cooperation does not rely on other players coordinating on a punishment trajectory following deviation by a given player. Instead, the deviator punishes itself by foregoing the benefits of the future incremental innovations that its contribution would have generated among its competitors if it had been made public. Our theoretical analysis sheds light on the pervasiveness of “open innovation”, as theorized by Chesbrough, and predicts that open innovation is enhanced when the number of innovative firms in the relevant industry goes up.

Further research could focus on more specific aspects. For example, in Silicon Valley, worker mobility has been an important vector of innovation sharing. It could be valuable to further analyze the role of the labor market in the diffusion of knowledge, which has been a largely untouched topic until now.¹⁶ In principle, such research avenues could deliver insights as to why Silicon Valley emerged in an economy with a flexible labor market as opposed to a European country where collective bargaining and labor regulation hamper

labor mobility. Another potential avenue of further research would be the analysis of an incomplete information network in which firms would release innovations only to a subset of firms to which it is ‘connected’: the incentive for a firm to make its innovations public would then depend on its position in a network.

Funding: EUR Grant ANR-17-EURE-0001.

Data Availability Statement: Data are contained within the article.

Acknowledgments: This paper has originally been circulated as a working paper under the title “Information sharing and cumulative innovation in business networks”. I am grateful to one anonymous referee for suggesting an improvement in the title. The paper has benefitted from comments from participants to the Toulouse lunch seminar, the MIT Macroeconomics seminar, and the Stanford Institute of Theoretical Economics; I am especially grateful to Xavier Gabaix, Jean-Charles Rochet and Jean Tirole. I am also grateful to five anonymous referees from this journal.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Proposition A1. Assume λ remains between bounds $\underline{\lambda}$ and $\bar{\lambda}$, such that $c(\underline{\lambda}) = 0, c(\bar{\lambda}) = +\infty$. Let

$$\alpha = \bar{f} \left[\frac{1}{r + N\underline{\lambda}} + \frac{(N-1)\bar{\lambda}}{(r + (N-1)\bar{\lambda} + \underline{\lambda})(r - \bar{\lambda}N(A-1))} \right];$$

$$\beta = \frac{\bar{\lambda}}{r + (N-1)\underline{\lambda} + \bar{\lambda}}.$$

Assume the following inequality holds:

$$\alpha + \beta \frac{A\bar{f}}{r - \bar{\lambda}N(A-1)} \leq \frac{1 - c(\lambda^*)}{r - \lambda^*N(A-1)},$$

then there exists a cooperative equilibrium where all firms constantly set $\lambda = \lambda^*$ and make all innovations public.

We follow the same steps as in the proof of Proposition 1. At each conceivable point in time, there exists a vector of $\vec{n} = (n_1, \dots, n_N)$ of technological levels and a vector $\vec{\lambda} = (\lambda_1, \dots, \lambda_N)$ of R&D efforts. Assume we restrict λ_i to be a sole function of the state vector $\vec{n}$, $\lambda_i = \lambda_i(\vec{n})$. That is, we only consider “Markov strategies” that depend on the current vector of state variables. Clearly, if all agents consider that along any path Markov strategies are followed, it is indeed optimal for them to follow such a strategy. Then, the value of a firm i can be recursively written as

$$V_i(\vec{n}) = \frac{\pi_i(\vec{n})(1 - c(\lambda_i(\vec{n}))) + \sum_{j=1}^N \lambda_j(\vec{n}) [\max(V_i(T_j U_j \vec{n}), V_i(U_j \vec{n}))]}{r + \sum_{j=1}^N \lambda_j(\vec{n})}$$

Defining, as previously,

$$\bar{V}_i(n, p) = \max_{\substack{\vec{n}=(n_1, \dots, n_N) \\ n_i=n \\ n_j \leq p, j \neq i}} V_i(\vec{n}),$$

we again have the property that for any $\vec{n} = (n_1, \dots, n_N)$ such that $n_j \leq p < n_i = n, j \neq i$,

$$V_i(\vec{n}) \leq \frac{A^n \bar{f}(1 - c(\lambda_i(\vec{n}))) + \sum_{j \neq i} \lambda_j(\vec{n}) \bar{V}_i(n, p+1) + \lambda_i(\vec{n}), \bar{V}_i(n+1, n+1)}{r + \sum_{j=1}^N \lambda_j(\vec{n})}. \quad (A1)$$

We can again prove that for $n = \max(n_i)$,

$$V_i(\vec{n}) \leq \frac{A^n \bar{f}}{r - \bar{\lambda}N(A-1)},$$

along the same lines as in the proof of Proposition 1. An implication is that for $p \leq n$

$$V_i(n, p) \leq \frac{A^n \bar{f}}{r - \bar{\lambda}N(A-1)} \quad (\text{A2})$$

Plugging into (A1) yields

$$V_i(\vec{n}) \leq \frac{A^n \bar{f}(1 - c(\lambda_i(\vec{n}))) + \left(\sum_{j \neq i} \lambda_j(\vec{n})\right) \frac{A^n \bar{f}}{r - \bar{\lambda}N(A-1)} + \lambda_i(\vec{n}) \frac{A^n \bar{f}}{r - \bar{\lambda}N(A-1)}}{r + \sum_{j=1}^N \lambda_j(\vec{n})}. \quad (\text{A3})$$

Then, observe that the following inequalities hold:

$$\frac{\sum_{j \neq i} \lambda_j(\vec{n})}{r + \sum_{j=1}^N \lambda_j(\vec{n})} \leq \frac{(N-1)\bar{\lambda}}{r + (N-1)\bar{\lambda} + \underline{\lambda}},$$

$$\frac{\lambda_i(\vec{n})}{r + \sum_{j=1}^N \lambda_j(\vec{n})} \leq \frac{\bar{\lambda}}{r + (N-1)\bar{\lambda} + \bar{\lambda}}.$$

Therefore, we have that

$$V_i(\vec{n}) \leq \frac{A^n \bar{f}}{r + N\underline{\lambda}} + \frac{(N-1)\bar{\lambda}}{r + (N-1)\bar{\lambda} + \underline{\lambda}} \frac{A^n \bar{f}}{r - \bar{\lambda}N(A-1)} + \frac{\bar{\lambda}}{r + (N-1)\bar{\lambda} + \bar{\lambda}} \frac{A^{n+1} \bar{f}}{r - \bar{\lambda}N(A-1)}. \quad (\text{A4})$$

Since this holds for all vectors such that $n_j \leq p < n_i = n$, $j \neq i$, we have, for $p < n$:

$$\begin{aligned} \bar{V}_i(n, p) &\leq \frac{A^n \bar{f}}{r + N\underline{\lambda}} + \frac{(N-1)\bar{\lambda}}{r + (N-1)\bar{\lambda} + \underline{\lambda}} \frac{A^n \bar{f}}{r - \bar{\lambda}N(A-1)} \\ &\quad + \frac{\bar{\lambda}}{r + (N-1)\bar{\lambda} + \bar{\lambda}} \frac{A^{n+1} \bar{f}}{r - \bar{\lambda}N(A-1)} \\ &= A^n \left(\alpha + \beta \frac{A \bar{f}}{r - \bar{\lambda}N(A-1)} \right) \end{aligned} \quad (\text{A5})$$

where

$$\alpha = \bar{f} \left[\frac{1}{r + N\underline{\lambda}} + \frac{(N-1)\bar{\lambda}}{(r + (N-1)\bar{\lambda} + \underline{\lambda})(r - \bar{\lambda}N(A-1))} \right];$$

$$\beta = \frac{\bar{\lambda}}{r + (N-1)\bar{\lambda} + \bar{\lambda}}.$$

Note that if we can prove that $\bar{V}_i(n, n-1) \leq V_C(n)$ regardless of the strategies followed in a deviation, then clearly there exists a cooperative equilibrium. Given (A5) and that

$$V_C(n) = \frac{A^n(1 - c(\lambda^*))}{r - \lambda^*N(A-1)},$$

a sufficient condition is

$$\alpha + \beta \frac{A\bar{f}}{r - \bar{\lambda}N(A-1)} \leq \frac{1 - c(\lambda^*)}{r - \lambda^*N(A-1)}, \quad (\text{A6})$$

where λ^* is the equilibrium value of λ in a cooperative equilibrium, that is any solution to (17).

The condition in Proposition A1 is non-empty, since it collapses to that in Proposition 1 for $\bar{\lambda} = \underline{\lambda} = \lambda$ and $c(\cdot) = 0$. By continuity, one can then construct examples where both sides of that inequality are arbitrarily close to the corresponding expressions in Proposition 1.

Notes

- ¹ See also Curley and Salmelin (2017) for a survey of the innovative culture of the software industry.
- ² See Nonaka and Nishiguchi (2001); Nonaka et al. (2001), and, in that same volume, the case studies of Nishiguchi (2001); Nishiguchi and Caspary (2001); Lincoln and Ahmadjian (2000), and de Michelis (2001).
- ³ See, for example, Carvalho Vieira et al. (2021); Hitchen et al. (2017); Rexhepi et al. (2019), and Seyfettinoglu et al. (2020).
- ⁴ On the other hand, obviously, many firms avoid making their innovations open and rely instead on patents or even trade secrets. A famous example is the Coca-Cola beverage formula (see Crittenden et al. 2015). Clearly, revealing such a secret would be of little interest to the firm, as it is likely to lead to imitation instead of cumulative process innovation (see Morikawa 2014 for a discussion of the effect of trade secrets on product vs. process innovation). Another famous example is the Google search algorithm. Here, the explanation is likely to be different. While revealing that algorithm is likely to lead to a wave of improvements that Google could benefit from, provided they were made public, the firm's market share is so big that it loses too much in the short term by making its algorithm public. In terms of the model, this is similar to a case where the number of firms is too small for the cooperative equilibrium to exist.
- ⁵ See Tirole (1990) for a survey.
- ⁶ This is the option in Aloysius (1999). See also Brod and Shivakumar (1997). Petit and Tolwinski (1999) assume away any problem in enforcing cooperation.
- ⁷ An interesting paper by Dutta and Seabright (2002) looks at the impact of the extent to which knowledge is explicit on growth and at its cross-impact with competition. However, the degree of explicitness of knowledge is exogenous, whereas it can be viewed as endogenous in the current paper. Katz and Ordover (1990) discuss informally how improvements in intellectual property rights may enhance incentives to share information by licensing. Licensing is not considered in the present paper.
- ⁸ We can simply assume that firms do observe whether one competitor is struck by an innovation even though they need not benefit from the innovation. Thus, there is no asymmetry of information as far as the structure of the game is concerned. The asymmetry of information with respect to the contents of the innovation are summarized by the payoffs and play no further role in the determination of the equilibrium strategies.
- ⁹ One has to check that V_i exists, i.e., is finite. That is actually an implication of the derivations that follow.
- ¹⁰ In the model's setting, a deviation would be detected as other firms would observe a fall in profits without an innovation being released, but that would change if profits were subject to shocks.
- ¹¹ This restriction implies that $A^{-k}f(k)$ goes up with k , so that $\lim_{k \rightarrow +\infty} A^{-k}f(k) = \sup A^{-k}f(k)$.
- ¹² It is slightly weaker than (4).
- ¹³ Note that we rule out any cooperation on λ , which may be unobservable, and continue to focus on the incentive to make one's innovation public.
- ¹⁴ Furthermore, this restriction prevents degenerate solutions where the growth rate is greater than the interest rate.
- ¹⁵ The model is clearly silent about consumer surplus. One may conceive of innovations that reduce the elasticity of demand and may thus harm consumers.
- ¹⁶ Lerner and Tirole (2002), though, deal with the role of career concerns in the development of open source software.

References

- Aghion, Philippe, and Peter Howitt. 1992. A Model of Growth Through Creative Destruction. *Econometrica* 60: 323–51. [CrossRef]
- Aloysius, John A. 1999. Membership in a Research Consortium: The project selection game. *Journal of Economic Behavior and Organization* 40: 325–36. [CrossRef]
- Baumol, William. 1992. Horizontal collusion and innovation. *Economic Journal* 102: 129–37. [CrossRef]

- Bessen, James, and Eric Maskin. 2008. Sequential Innovation, Patents, and Imitation. *The Rand Journal of Economics* 40: 611–35. [CrossRef]
- Brod, Andrew, and Ram Shivakumar. 1997. R&D cooperation and the joint exploitation of R&D. *Canadian Journal of Economics* 30: 673–84.
- Cabigiosu, Anna. 2023. Open and collaborative innovation in the automotive industry. In *The Green Transition of the Automotive Industry*. Edited by A. Cabigiosu and P. Lancini. Cham: Springer, pp. 131–59.
- Carvalho Vieira, Filipe, Leandro R. C. Bonfim, and Aline Cristina da Cruz. 2021. The process of opening innovation networks: Open innovation at Embrapa Florestas. *Innovation & Management Review* 19: 123–39.
- Chesbrough, Henry W. 2003. *Open Innovation. The New Imperative for Creating and Profiting from Technology*. Harvard: Harvard Business School Press.
- Chesbrough, Henry W., Wim Vanhaverbeke, and Joel West, eds. 2014. *New Frontiers in Open Innovation*. Oxford: Oxford University Press.
- Cooper, Russell, and Andrew John. 1988. Multiple equilibria and strategic complementarities. *Quarterly Journal of Economics* 103: 441–63. [CrossRef]
- Cozzi, Guido. 1999. R and D cooperation and growth. *Journal of Economic Theory* 86: 17–49. [CrossRef]
- Crittenden, William F., Victoria L. Crittenden, and Allison Pierpont. 2015. Trade secrets: Managerial guidance for competitive advantage. *Business Horizons* 58: 607–13. [CrossRef]
- Curley, Martin, and Bror Salmelin. 2017. *Open Innovation 2.0. The New Mode of Digital Innovation for Prosperity and Sustainability*. Berlin/Heidelberg: Springer. Available online: <https://link.springer.com/book/10.1007/978-3-319-62878-3> (accessed on 23 January 2024).
- d’Aspremont, Claude, and Alexis Jacquemin. 1988. Cooperative and Noncooperative R&D in Duopoly with spillovers. *American Economic Review* 78: 1133–37.
- de Michelis, Giorgio. 2001. Cooperation and Knowledge Creation. In *Knowledge Emergence: Social, Technical, and Evolutionary Dimensions of Knowledge Creation*. Edited by Ikujiro Nonaka and Toshihiro Nishiguchi. Oxford: Oxford University Press, pp. 124–46.
- Dutta, Jayasri, and Paul Seabright. 2002. Competition and Economic Growth. Available online: https://www.researchgate.net/publication/255653065-Competition_and_Economic_Growth (accessed on 23 January 2024).
- Fudenberg, Drew, and Jean Tirole. 1991. *Game Theory*. Cambridge: MIT Press.
- Grossman, Gene M., and Elhanan Helpman. 1993. *Innovation and Growth in the Global Economy*. Cambridge: MIT Press.
- Hitchen, Emma L., Petra A. Nylund, and Eric Viardot. 2017. The Effectiveness of Open Innovation: Do Size and Performance of Open Innovation Groups Matter? *International Journal of Innovation Management* 21: 1–24. [CrossRef]
- Hyde, Alan. 1998. The Wealth of Shared Information: Silicon Valley’s High-Velocity Labor Market, Endogenous Economic Growth, and the Law of Trade Secrets. mimeo, Rutgers U. Available online: <http://andromeda.rutgers.edu/~hyde/> (accessed on 23 January 2024).
- Katz, Michael, and Janusz Ordover. 1990. R&D cooperation and competition. In *Brookings Papers on Economic Activity*. Washington, DC: Brookings Papers on Economic Activity.
- Leahy, Dermot, and Peter Neary. 1997. Public policy toward R&D in oligopolistic industries. *American Economic Review* 8: 642–62.
- Lerner, Josh, and Jean Tirole. 2002. *Some Economics of Open Source* Mimeo. Cambridge: Harvard and Toulouse University.
- Lincoln, James, and Christina Ahmadjian. 2000. Shukko (Employee Transfers) and Tacit Knowledge Exchange in Japanese Supply Networks: The Electronics Industry Case. In *Knowledge Emergence: Social, Technical, and Evolutionary Dimensions of Knowledge Creation*. Edited by Ikujiro Nonaka and Toshihiro Nishiguchi. Oxford: Oxford University Press, pp. 247–69.
- Morikawa, Masayuki. 2014. *Innovation in the Service Sector and the Role of Patents and Trade Secrets*. AJRC Working Papers 1403. Canberra: Australia-Japan Research Centre, Crawford School of Public Policy, The Australian National University.
- Nishiguchi, Toshihiro. 2001. Coevolution of Interorganizational Relations. In *Knowledge Emergence: Social, Technical, and Evolutionary Dimensions of Knowledge Creation*. Edited by Ikujiro Nonaka and Toshihiro Nishiguchi. Oxford: Oxford University Press, pp. 197–222.
- Nishiguchi, Toshihiro, and Sigrun Caspary. 2001. “Co-opetition” in the Japanese Aircraft Industry. In *Knowledge Emergence: Social, Technical, and Evolutionary Dimensions of Knowledge Creation*. Edited by Ikujiro Nonaka and Toshihiro Nishiguchi. Oxford: Oxford University Press, pp. 223–46.
- Nonaka, Ikujiro, and Toshihiro Nishiguchi. 2001. Introduction: Knowledge Emergence. In *Knowledge Emergence: Social, Technical, and Evolutionary Dimensions of Knowledge Creation*. Edited by Ikujiro Nonaka and Toshihiro Nishiguchi. Oxford: Oxford University Press, pp. 3–12.
- Nonaka, Ikujiro, Noboru Konno, and Ryoko Toyama. 2001. Emergence of “Ba”: A Conceptual Framework for the Continuous and Self-transcending Process of Knowledge Creation. In *Knowledge Emergence: Social, Technical, and Evolutionary Dimensions of Knowledge Creation*. Edited by Ikujiro Nonaka and Toshihiro Nishiguchi. Oxford: Oxford University Press, pp. 13–29.
- Petit, Maria Luisa, and Boleslaw Tolwinski. 1999. R&D cooperation or competition? *European Economic Review* 43: 185–208.
- Piore, Michael, and Charles Sabel. 1986. *The Second Industrial Divide: Possibilities for Prosperity*. New York: Basic Books.
- Polanyi, Michael. 1962. *Personal Knowledge: Towards a Post-Critical Philosophy*. Chicago: University of Chicago Press.
- Purificato, Marco. 2014. The Open Innovation Paradigm in Electric Vehicle Industry: A Case Study of Tesla Motors. Bachelor’s thesis, LUISS Guido Carli, Rome, Italy.

- Raymond, Eric S. 1999. The Cathedral and the Bazaar. Available online: https://monoskop.org/images/e/e0/Raymond_Eric_S_The_Cathedral_and_the_Bazaar_rev_ed.pdf (accessed on 23 January 2024).
- Rexhepi, Gadaf, Robert D. Hisrich, and Veland Ramadani, eds. 2019. *Open Innovation and Entrepreneurship*. Springer Books. Berlin/Heidelberg: Springer.
- Romer, Paul M. 1990. Endogenous Technological Change. *Journal of Political Economy* 98: S71–S102. [CrossRef]
- Saint-Paul, Gilles. 2003. Growth effects of non proprietary innovation. *Journal of the European Economic Association* 1: 429–39. [CrossRef]
- Schumpeter, Joseph A. 1942. *Capitalism, Socialism, and Democracy*. London: Routledge.
- Scotchmer, Suzanne. 1991. Standing on the Shoulders of Giants: Cumulative Research and the Patent Law. *Journal of Economic Perspectives* 5: 29–41. [CrossRef]
- Scotchmer, Suzanne. 1996. Protecting Early Innovators: Should Second-Generation Products be Patentable? *RAND Journal of Economics* 27: 322–31. [CrossRef]
- Scotchmer, Suzanne. 2004. *Innovation and Incentives*. Cambridge: MIT Press.
- Seyfettinoglu, Ümit K., Şebnem Arik, and Selim Çagatay. 2020. Determinants of Industry and Region Based Open Innovation in Turkey. *Istanbul Business Research* 49: 1–35. [CrossRef]
- Tirole, Jean. 1990. *The Theory of Industrial Organization*. Cambridge: MIT Press.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

Journal of Risk and Financial Management Editorial Office

E-mail: jrfm@mdpi.com
www.mdpi.com/journal/jrfm



Disclaimer/Publisher's Note: The title and front matter of this reprint are at the discretion of the Guest Editor. The publisher is not responsible for their content or any associated concerns. The statements, opinions and data contained in all individual articles are solely those of the individual Editor and contributors and not of MDPI. MDPI disclaims responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-3414-3