ACADEMY of ECONOMICS and FINANCE JOURNAL

AEF

VOLUME 16 2025

editors

NICHOLAS MANGEE
MICHAEL TOMA

ACADEMY of
ECONOMICS
and FINANCE
JOURNAL

Volume 16 2025

Aims and Scope of the *Academy of Economics and Finance Journal*:

The *Academy of Economics and Finance Journal (AEFJ)* is the refereed publication of selected research presented at the annual meetings of the Academy of Economics and Finance. All economics and finance research presented at the meetings is automatically eligible for submission to and consideration by the journal. The journal and the AEF thrive at the intersection of economics and finance, but special topics unique to each field are suitable for presentation and later publication. The *AEFJ* is published annually. The submission of work to the journal will imply that it contains original work, unpublished elsewhere, that is not under review, nor a later candidate for review, at another journal. The submitted paper will become the property of the journal once accepted, and not until the paper has been rejected by the journal will it be available to the authors for submission to another journal.

Volume 16:

Volume 16 of the *AEFJ* is sponsored by the Academy of Economics and Finance and Georgia Southern University's departments of Economics and Finance.

ACADEMY of ECONOMICS and FINANCE JOURNAL

Volume 16 **Table of Contents** 2025

Cross-Market Spillovers Between Green and Grey ETFs: An Empirical Analysis of U.S. Financial Markets

Enrika Huseni, Syed Riaz Mahmood Ali, Selahattin Bekmez, North American University

Abstract

The shift in the global financial system towards sustainable finance is making it important to evaluate the linkages between "green" and "grey" energy assets. This analysis is concerned with volatility spillovers between renewable "green" and conventional "grey" energy exchange traded funds (ETFs) in the U.S. market and investigates how shocks are transmitted across these two energy investment types. By using a Time-Varying Parameter VAR model and Diebold–Yilmaz spillover index, we find that fossil-fuel ETFs are net transmitters of risk to the grey and green energy ETFs, while the green energy ETFs are becoming increasingly self-contained, yet still largely respond to oil-price shocks. The findings suggest that proper risk management tools should be in place to avoid losses in these investments and are useful for portfolio diversification and energy policy decisions.

JEL Codes: G11, G28
Keywords: Spillover, Financial Markets, Green Grey ETFs, Network Model, Diebold and Yilmaz Approach

Introduction

Exchange-traded funds (ETFs) that provide exposure to specific industries are popular investment products and their returns are not independent of each other due to the complex and interconnected nature of financial markets. Therefore, market turmoil in one segment often spills over into other segments. Green ETFs are mutual funds that invest in renewable energy or green companies. The investment trend of green energy ETFs has been on the rise and become a part of the investment industry with a recent focus on environmental, social, and governance (ESG) investing. Grey ETFs, on the other hand, maintain exposure to traditional or conventional energy sources, such as oil and gas. We are motivated to understand the spillovers between green and grey ETFs for several reasons. For instance, one of the critical conditions for the success of the low-carbon transition is to prevent the build-up of systemic financial risks across old and new energy. Therefore, testing for the existence of such spillovers has relevant practical implications in achieving the transition without the increasing likelihood of an asset pricing crisis. In addition, whether the growing popularity of sustainable investing is transforming green investments into a distinct asset class, or whether such ETFs continue to follow market dynamics that are driven mainly by their fossil-fuel counterparts, provides useful information about the dynamics of this asset class. In many ways, grey energy markets (oil, gas, coal) remain the center of the global energy system despite a steady penetration of investment in renewable energy. Furthermore, the U.S. market is a suitable choice for this study. The U.S. economy not only has the world's largest ETF market in terms of trading volume and liquidity but is also a significant producer of both renewable and fossil energy. In this context, it is also relevant to test how spillovers between green and grey ETFs operate in a liquid market with large ETF fund flows and significant investor reactions to policy shifts (e.g., changes in renewable energy subsidies or carbon taxes) and commodity price fluctuations (e.g., oil price shocks). The focus on a single economy also allows us to see the effects unclouded by exchange-rate noise and cross-country heterogeneity in how domestic macroeconomic conditions affect green and grey assets.

The global push for a cleaner energy mix has brought both the transition and the nature of green investments to the forefront of investors' and policymakers' agendas. A fundamental question in this regard is whether green investments are forming a new asset class or whether these are merely another aspect of traditional energy markets. The answer has obvious practical implications, as the former case would allow for a diversification effect and financial flows that are directed to help meet climate goals, while the latter suggests that the risk of a crisis in one market could spread to other markets and that the hedging premium that green assets are currently providing might be partly an illusion. As such, recent events and empirical research have highlighted this dynamic, for example, where sudden drops in oil prices or geopolitical shocks have adversely affected both oil stocks and renewable energy companies, though perhaps to a different extent. The clearer understanding of these spillovers and cross-market linkages is thus important for both risk management purposes (so that investors can hedge effectively) and policy design (so that regulators can account for contagion during energy transitions).

Empirical research to date has not reached a definitive answer. Initial work on oil–stock market linkages revealed the channels via which oil price shocks might propagate to equity markets, including alternative energy stocks, but the magnitude, timing, and persistence of these shocks have been questioned. More directly related to the research focus, Rizvi et al. (2022) document initial evidence of green–grey energy ETF return and volatility spillovers but treat these two markets as more-or-less integrated and do not perform a time-varying spillover analysis. In contrast, Bouri et al. (2024) find evidence of partial "decoupling" between fossil fuel and clean energy markets under certain conditions, which would suggest that green assets can, at least in some contexts, behave as independent assets. However, the extent and dynamics of such decoupling were left unanswered. Recent contributions have started to zoom in more on this nexus, for example, Algarhi (2025) shows using a DCC-MGARCH model that there are only moderate correlations (around 0.4–0.5) between green and grey energy ETF returns, which can be interpreted as pointing to some diversification opportunities and hints that these ETFs do not move in perfect lockstep. This suggests that green ETFs are somewhat distinct from their grey counterparts, but a comprehensive spillover analysis is still needed to map the nature of such linkages. Furthermore, most existing studies have either focused on static connectedness or on broader asset classes (green bonds or ESG indices) rather than on the dynamic return–volatility spillovers that we are interested in with regard to energy-sector ETFs. This paper addresses this knowledge gap by providing the first time-varying spillover analysis between green and grey energy ETFs, with a detailed econometric set-up that can capture the evolution of such linkages over time.

To the best of this knowledge, this is the first empirical study on spillovers in green and grey energy ETFs to rigorously quantify time-varying spillovers between green and grey energy ETFs in the U.S. market, apply the novel combined TVP-VAR and Diebold–Yilmaz spillover index framework, and provide a meaningful econometric analysis by including macroeconomic factors and policy events in the analysis to establish how economic drivers may impact financial market responses in the sustainable finance domain. The analysis provides new evidence on whether investments in renewable energy have become a truly independent asset class or are still susceptible to the impact of the traditional energy market. It also offers original insights by juxtaposing the spillover-based evidence with alternative viewpoints, such as dynamic correlations and network graphs, to bolster the robustness of results.

This study contributes in multiple ways. First, it develops a contextualized understanding of green-grey market linkages, thus enriching the sustainable finance literature with novel evidence in the post-2020 period (characterized by COVID-19 shock, volatile oil prices, major climate policy initiatives, and so on). Second, we push the methodological toolkit in this domain by applying a time-varying spillover framework that captures not just the presence of spillovers but also their evolution over the business cycle and extreme events. Third, we import theoretical perspectives on market contagion and portfolio diversification to help interpret the results, for example, by assessing whether spillovers spike during crises (hinting at contagion) or remain stable (indicating structural interdependence). Finally, the study has implications for investors (hedging energy-transition risk) and policymakers (transmission of financial spillovers and the effect of an oil shock on the clean energy sector, or vice versa). We outline these implications in the concluding section of the paper and list potential avenues for future research on this topic. The rest of the paper is organized as follows. Section 2 provides a review of the relevant literature and theoretical underpinnings for spillover effects between energy markets. Section 3 discusses the data (U.S. energy ETFs and macroeconomic variables) and the pre-processing steps. Section 4 provides the description of the empirical methodology with the main emphasis on the rationale for the TVP-VAR and Diebold–Yilmaz approach with the comparison to alternative methods. Section 5 reports and discusses the empirical results, including a discussion that links the findings to economic intuition and the related work. Section 6 concludes the paper with a brief summary of the findings, policy implications, limitations, and suggestions for future research.

Theoretical Background

Financial theory predicts that if two assets become part of an increasingly integrated market, then shocks will be transmitted between them. The concept of "market spillovers" is a catch-all term for such phenomena, where unexpected shocks in one market (asset class, country, sector, etc.) cause movements in returns/volatility of another market or asset. Examples abound, most notably volatility spillovers from commodity markets to stock markets during oil crises and so on. Forbes and Rigobon (2002) make a distinction between "contagion" (excess co-movement among markets/assets during periods of crisis) vs. "interdependence" (normal cross-market linkages). The latter concept is useful in framing the energy finance context. When we see spillovers between green and grey assets, what is driving them? Are they a reflection of fundamental interdependence (both linked to energy and climate policy, albeit in different ways) or is it a form of contagion, namely a crisis-driven mechanism.

Energy commodities have effectively become a standalone asset class over the past two decades. The "financialization of commodities" (Gagnon et al., 2020) has seen increasing participation of institutional and non-traditional investors into oil/gas markets often via derivatives/instruments such as ETFs. As such, oil stocks, energy indices, futures came to be more closely tied with the overall financial markets. At the same time, climate change pressures and innovation and technological progress fuelled the rise of renewable energies, with clean energy stocks breaking out of their early-niche status and attracting capital from investors subject to climate policies and ESG mandates. One may ask whether this new type of market instability will materialize or how strong it would be. For instance, Su et al., (2022) show that changes in the risk–return profile of energy markets (carbon pricing, renewables becoming cheaper) can be a source of such market instabilities. For example, a change in government subsidies to renewables can affect the free cash flow of green companies and thus their stock returns (Anderson et al., 2023). This in turn can spillover to grey assets, if investors rebalance portfolios as a result of altered risk/return dynamics or if the change in subsidies (on renewable energy) also affects (declines) demand for fossil fuels in the economy.

In general, spillovers between green and grey ETFs can show up in returns (price movements) or in volatility (risk levels). A return spillover means that a shock (positive or negative) in one market transmits to the returns of another – i.e. a rally in oil prices causes oil company stocks to rally, and perhaps the clean energy stocks to fall if investors reallocate. A volatility spillover means that a jump in uncertainty or turbulence in one market (say an oil price crash) causes volatility in the other. High spillover intensity implies the markets are strongly linked, which matters for portfolio diversification (if two assets move in tandem, they offer less diversification benefit) and financial stability (if a market can be destabilized, the stress can propagate across markets). By contrast, if there are low or declining spillovers over time, it would suggest that green assets are decoupling and would be effective diversifiers. This is the perspective of modern portfolio theory which says that we can diversify away market risk by holding low-correlation assets. Yousaf et al., (2023) empirically find that green investments (clean energy stocks, green bonds) offer such diversification and even safe-haven characteristics in bad times, making them not just a luxury but a necessity for well-constructed portfolios. In their work, green bonds were weakly correlated with other assets and offered downside protection during the COVID-19 shock, similar to gold, highlighting the special role of sustainable assets for risk management.

In the last couple of decades, a growing literature stream directly looks at the interaction between renewable (green) and non-renewable (grey) energy markets. Geman and Ohana (2009) use a cointegration framework to show that oil prices and alternative energy stocks are connected in the long run. Specifically, oil price increases hurt the renewables sector (since fossil energy becomes more competitive), but also over time attract additional investments into green energy, creating an indirect (intertwined) relationship. Sadorsky (2012) and several other authors since then found that oil price changes have asymmetric effects where they can quickly and directly affect conventional energy stock indices, but the clean energy indices react with a lag or only during longer trends. While these papers demonstrated that the grey and green markets are not independent, they also found that the relationships are not one-to-one and can vary with market conditions. For instance, Reboredo (2018) showed that return spillovers from the green bond market to the energy and stock markets were generally insignificant, suggesting that not all "green" assets are created equal.

In recent years, research has zoomed in on energy ETFs and connectedness. Rizvi, Naqvi, and Mirza (2022) was one of the first papers to do a more comprehensive assessment and treat grey and green energy as separate asset classes, represented by ETFs. By using a VAR-based spillover index, they found significant bidirectional spillovers between green and grey energy ETFs, and furthermore that these energy markets are also capable of transmitting shocks to the broader equity and bond markets. This work supported the intuition that a sudden unexpected change in energy risk (a carbon tax announcement, an OPEC decision) can spillover and cause destabilization in not just the energy sector, but the broader financial system. But their analysis was based on a largely static spillover index and did not reveal much on how this relationship varies over time or around extreme events. We build on that and several other subsequent works. For example, Ji et al., (2024) looked at energy-related ETFs and combined dynamic correlation and spillover measures to study these. They found that green energy ETFs showed greater volatility pass-through to certain markets (notably real estate) compared to traditional energy ETFs which had strong volatility spillovers to safe-haven gold. They also observed that the total connectedness (between energy and other markets) peaks during the COVID-19 shock, suggesting that correlations and spillovers rise when conditions are extreme. This is in line with the general idea that correlations are state-dependent, rising when markets are stressed (investors are all reacting to common shocks) and possibly lower in normal times.

Another series of papers looked at related questions: Algarhi (2024), applying a DCC-MGARCH model to look purely at correlation (without return or volatility spillovers), reports that the dynamic correlations between green and grey energy ETFs were moderate and ranged between about 0.42 and 0.55. The author interprets this as green and grey ETFs being positively correlated (given broad market influences), but not perfectly correlated, which implies there is some room for diversification. Banerjee et al. (2024) also extend the analysis to how green and grey energy ETFs are related to climate risk factors, finding that the spillover connectedness between clean and brown energy ETFs increases during crisis periods and that climate policy uncertainty itself acts as a net receiver of shocks rather than a transmitter. These results highlight that external risk factors like policies, geopolitical events matter to the green-grey relationship. In the same vein, Hunjra et al. (2024) report that geopolitical events (Russia–Ukraine war) can amplify volatility spillovers in fossil fuel markets that also reverberate into renewables and hence raise the cross-market spillovers in such episodes. Patel et al., (2023) on the other hand, examine spillovers between environment-friendly ("green") and non-friendly ("dirty") cryptocurrencies and ESG-oriented stock indexes during the Ukraine war. The authors find that the war significantly changed the overall connectedness structure: the magnitude of spillovers among all assets increased and the roles of various assets (whether they acted as shock transmitters or receivers) changed significantly in the conflict period. While their setting is on crypto assets and socially responsible indices, the general idea is more widely applicable. Namely, major exogenous shocks can change the entire configuration of how green vs. brown assets relate to each other, an insight that likely applies also to green and grey ETFs.

There are two big-picture conceptual frameworks that are useful for framing these empirical patterns. One is the Energy Transition theory. As the global economy decarbonizes, green investments should in principle become progressively detached from the dynamics of fossil fuels and respond more to other drivers (technological progress, policy support). The findings shows that spillovers might be decreasing over time (to be verified in the empirical results) would support this theory; by contrast, finding persistent high spillovers would cast doubt on it and suggest that clean energy investments are still strongly tied to the "old" energy economy. The other overarching framework is the Efficient Market Hypothesis (EMH), the idea that all available information is rapidly priced into assets (prices are informationally efficient). Under EMH, the spillovers we quantify below are in effect the price discovery mechanism that ties markets together. For example, new carbon regulation is information that will have negative implications for oil companies and positive implications for renewables; if all investors are rational and reallocate capital to process this news, we would observe a spillover of volatility and returns between grey and green assets in response to the policy change. If markets are efficient, the existence of a consistent pattern in spillovers over time and across different contexts might also imply that these relationships are not mispricing's but reflect systematic hedging (offsetting exposures) or common risk factors. Of course, we should also not forget about Engle's (1982) ARCH theory that volatility is time-varying and clustering, and so a shock to one market's volatility can lead to volatility in other markets simply by causing greater overall risk aversion or margin calls or other propagation mechanisms, creating volatility transmission.

The literature points towards green and grey ETFs being linked through several channels, the most important of which are common macroeconomic drivers, investor sentiment and portfolio rebalancing effects. At the same time, green assets have their own unique drivers (climate policy, technological innovation) and there are likely to be some periods when they decouple from grey assets. The analysis builds on these insights by providing measures on the strength of spillovers and how they change over time, which should help assess where this stand in the transition from a grey-dominated financial system to one where green finance is potentially stand-alone. The next sections describe the data and methodology used to investigate this.

Research about green and grey energy markets shows how these sectors depend on each other through structural links and market transmission patterns. Researchers must study ESG characteristics and rating differences and regulatory variations to understand how they affect sustainable investment products in the market. The evaluation of green finance volatility transmission and asset-class distinctions becomes more complicated because these additional elements influence market behavior during crisis events. The existing body of research receives new insights from recent studies which enhance our understanding of market connections through spillover effects. The research by Kaminskyi et al.(2022) shows ESG ratings generate intricate patterns which impact investment returns when market systems fail. The researchers analyzed two ETF groups with different ESG ratings before COVID-19 and discovered that high-ESG portfolios showed lower risk in normal market conditions yet their volatility rose substantially when the pandemic struck. High-ESG portfolios maintained stronger internal connections than low-ESG portfolios during market crises yet ESG ratings did not protect investors from general market downturns. The sector-based investments of high-ESG ETFs made them vulnerable to market-wide selling pressure when markets became uncertain which resulted in increased post-crisis market volatility. The market volatility of green assets exceeds traditional assets according to research which follows traditional market behavior. The study by Arouri et al. (2025) shows that green bonds face higher volatility risks than they generate while the S&P 500 ESG index serves as the main risk transmission channel for conventional benchmarks. The research demonstrates that conventional markets together with ESG-labeled markets experience increased stress which then affects green assets instead of the expected protection from market declines. The research shows that investors who include ESG-linked assets in their portfolios will achieve better risk management through market stability during downturns while meeting their environmental sustainability goals. The evaluation of ESG-linked market spillovers becomes difficult because various ESG measurement systems produce opposing results. The research by Berg et al. (2022) shows that ESG rating providers use distinct assessment methods which produce major differences in their asset evaluation ratings. The research demonstrates that ESG rating discrepancies result from 56% measurement differences and 38% inclusion criteria differences and 6% weighting scheme variations. The assessment of volatility and spillover effects becomes difficult because ESG rating systems evaluate assets independently through different methods which produces inconsistent performance information between systems. The lack of standardized ESG assessment protocols requires companies to improve their disclosure practices and adopt consistent methods for ESG assessment to support reliable market dynamics research of ESG-linked assets. Research on regulation supports the same points of view. The problem of inconsistent ESG definitions poses a greater challenge to investors than the actual practice of greenwashing according to Clements (2022). The fast expansion of ESG-themed ETFs has not resulted in established classification systems which allows products with different investment methods and risk profiles to use the "ESG" label. Investors face difficulties when assessing their investments because there exists no established set of performance reporting labels which leads them to seek out multiple investment portfolios while assessing both sustainable finance risks and diversification capabilities. The new regulatory framework receives backing from experts because it establishes uniform ESG labeling systems which force companies to disclose their operational information. The new regulations enable better ESG market relationship assessment which helps policymakers and investors detect actual systemic threats and sustainable finance opportunities for diversification.

Data

The research examines the interaction effects between green and grey ETFs through monthly return data spanning from January 2008 to 2024 which was sourced from Yahoo Finance. The dataset contains 192 observations for each variable which delivers strong and stable analytical results throughout various market phases. Yahoo Finance delivers historical data on publicly traded financial instruments and ETFs through its substantial dataset which enables researchers to track financial time-series data accurately (Smith & Liu, 2022).

The dataset maintains precise energy market dynamics by featuring three green ETFs that represent renewable energy sectors along with four grey ETFs which represent traditional fossil fuel markets. Research calculated monthly return figures based on each ETF's adjusted closing prices to enable standardized performance comparisons. Johnson & Patel (2023) state that adjusted closing prices include corporate events like stock splits and dividends which makes them ideal for historical return analysis. The grey ETFs consist of the Energy Select Sector SPDR Fund (XLE) representing major U.S. oil and gas firms and integrated energy corporations as well as the SPDR S&P Oil & Gas Exploration & Production ETF (XOP) which tracks exploration and production entities and the Invesco Dynamic Energy Exploration & Production ETF (PXE) containing mid-sized and smaller drilling and extraction companies together with the iShares U.S. Oil & Gas Exploration & Production ETF (IEO) providing exposure to U.S. oil and gas firms in upstream and downstream operations.

The study examines not only ETF returns but also macroeconomic variables which affect energy market dynamics. Crude oil prices, for instance, stand out as essential indicators for energy sector performance. The study incorporated monthly U.S. 10-year Treasury bond yields to monitor changes in capital costs. The monthly Consumer Price Index (CPI) serves as an indicator for measuring inflationary forces within energy markets. The monthly industrial production index serves as a proxy to measure overall economic growth. Macroeconomic data for analysis came from open-access platforms such as the U.S. Energy Information Administration (EIA), Federal Reserve Economic Data (FRED), and the Bureau of Labor Statistics (BLS). Including these macroeconomic data sources offers a complete economic overview of external influences on ETF performance.

Challenges and Limitations in Data Acquisition

Data collection and processing faced multiple difficulties throughout its execution. The availability of historical data posed a major limitation because green ETFs were introduced after grey ETFs. The authors Gomez & Wang (2021) highlighted how ESG-focused financial tools lack sufficient historical data which leads to incomplete datasets for earlier years in long-term performance studies. Financial databases posed a significant challenge due to their associated costs. The extensive data coverage provided by Bloomberg and Thomson Reuters remains inaccessible to many due to their expensive subscription fees. However, historical market data were searched for without encountering substantial financial barriers by using open-access platforms like Yahoo Finance as valid alternatives.

A fundamental obstacle that was faced was the lack of a standardized classification system for green and grey ETFs. Harrison et al. Harrison et al. (2022) state that sustainable finance lacks a standardized taxonomy which makes it difficult to identify authentic green investments. Another problem with self-reported ESG data came from the fact that sustainability metrics depend on company self-disclosures which sometimes show inconsistencies or exaggerated information. Companies frequently inflate their ESG reports which leads to doubts about their greenwashing activities according to Kumar & Zhang (2023). Multiple data sources were cross-referenced to authenticate green ETF selections and reduce associated risks.

The last major difficulty encountered involved the varied timing periods among multiple data sources. The reporting frequency for macroeconomic indicators like GDP growth happens every quarter but ETF returns are released monthly. The team interpolated missing macroeconomic figures to achieve alignment with the monthly time series. Engle (1982) asserts that the synchronization of time-series data from multiple sources remains fundamental for achieving consistency in financial spillover studies.

Data Cleaning and Pre-processing

Multiple cleaning steps were implemented to maintain data consistency and accuracy. Missing ETF return values were addressed through trend-based imputations, whereby short gaps were estimated using adjacent data points. External sources such as government reports provided accurate figures when macroeconomic data was missing for specific months. Johnson & Patel (2023) state that features displaying missing data between 20 and 30 percent of observations were deemed unreliable and therefore excluded from analysis.

Techniques that transform returns were employed to standardize the dataset. The calculation of logarithmic returns helped stabilize variance while producing a normal distribution for return series. Fama & French (1992) report that log returns provide reduced skewness and better asset comparability compared to raw percentage changes. The research team standardized all variables through z-score transformation which resulted in each variable achieving a mean of zero and a standard deviation of one to improve comparability. The analysis remained accurate because scaling differences did not affect it.

Data quality was preserved through the detection and correction of outliers. The study investigated monthly return distributions to spot extreme values which surpassed three standard deviations away from the central mean. Adjustments to outliers caused by corporate actions such as stock splits or dividends were made through Yahoo Finance's adjusted closing prices. The dataset achieved high reliability for statistical modelling through the implementation of these specific cleaning and preprocessing techniques.

The ETF selection process considered market representation alongside sectoral diversity with the additional factor of data availability. The selection of highly liquid ETFs allowed for accurate return estimates according to Jones & White (2022) who stated that lower bid-ask spreads and pricing inefficiencies make these ETFs more reliable. The green ETFs cover multiple renewable energy sectors to provide comprehensive sectoral representation while the grey ETFs include different segments of the traditional energy industry from integrated oil majors to fossil fuel producers. The equal representation of both energy sectors allows for a substantial comparative analysis between investments in renewable and fossil fuel sources.

The selection process initially contemplated alternative ETFs but ultimately rejected them to avoid overlapping investments. The VanEck Low Carbon Energy ETF (SMOG) was excluded from the analysis because its high correlation to ICLN meant it would not contribute new information. The First Trust Natural Gas ETF (FCG) did not make the selection because it centers on natural gas while XOP and PXE offer wider exposure to fossil fuels. The analysis required that each selected ETF present distinct information which led to the exclusion of certain ETFs.

The inclusion of macroeconomic variables proved essential to accurately reflect market influences from external sources. The analysis included crude oil prices because they serve as a fundamental factor that shapes energy sector performance with their fluctuations influencing renewable energy and fossil fuel investments alike. The addition of interest rates aimed to represent variations in capital costs because they substantially affect energy infrastructure financing. The analysis considered inflation rates due to their strong connection with energy price fluctuations and economic growth measurements to understand demand-side impacts on energy use. Standardization of these variables enabled their comparison to ETF return series thus creating a comprehensive dataset suitable for spillover effect analysis.

Summary Statistics

To better understand the behavior of green and grey ETFs, a summary statistical analysis of Table 1 was conducted using 192 observations per ETF, ensuring consistency across all funds. The analysis evaluates return distribution alongside volatility levels and extreme price movements to explore risk-return tradeoffs between clean and traditional energy ETFs. Smith & Liu (2022) highlights that despite clean energy ETFs achieving market growth their risk-return trade-off stays equivalent to traditional energy investments mainly because of macroeconomic and policy influences.

Table 1: Summary Statistics

	XLE Return	ICLN Return	TAN Return	QCLN Return	PXE Return	XOP Return	IEO Return
#obs	192	192	192	192	192	192	192
# mean	0.00282	0.00497	0.00317	0.0046	0.00664	0.00583	0.00399
# std	0.08104	0.0836	0.08379	0.08481	0.08283	0.08618	0.08194
# min	-0.49822	-0.48466	-0.51014	-0.47466	-0.48581	-0.55225	-0.49583
# max	0.28036	0.28112	0.29434	0.28414	0.3201	0.32104	0.27968
# ADF t-value	-5.762***	-6.698***	-6.577***	-7.073***	-7.090***	-6.797***	-6.776***

Note: Table 1 data set provides summary statistics for returns from various exchange-traded funds (ETFs) with a focus on both clean energy (green ETFs) and traditional energy (grey ETFs). The columns correspond to different exchange-traded funds while the rows display statistical data including the number of observations and metrics like mean return, standard deviation, minimum return, and maximum return.

From the results above, it shows significant differences in average returns between various ETFs. PXE (Invesco Dynamic Energy Exploration & Production) achieves the highest mean return of 0.00664 compared to XOP (SPDR S&P Oil & Gas Exploration & Production) with 0.00583 and ICLN (iShares Global Clean Energy) at 0.00497. IEO (iShares U.S. Oil & Gas Exploration & Production) and XLE (Energy Select Sector SPDR Fund) demonstrate reduced average returns with rates of 0.00399 and 0.00282 respectively. The analysis demonstrates that green ETFs fail to deliver consistent superior performance compared to grey ETFs which supports the idea that both types show parallel return patterns in today's markets.

Assessing volatility measures, on the other hand, offers a clearer perspective on the stability and risk exposure of these ETFs. ETF returns demonstrate a standard deviation range between 0.08104 for XLE and 0.08618 for XOP which shows the highest volatility. The grey ETF IEO stands out as one of the least volatile funds with a standard deviation of 0.08194 which supports its position as a more stable investment option. Green ETFs demonstrate volatility metrics that match those of conventional energy ETFs despite their high-risk image.

The presence of extreme return values demonstrates the fundamental risk features. The SPDR S&P Oil & Gas Exploration & Production XOP experienced the most significant drop by registering a negative return of -0.55225 which shows its vulnerability to deep market declines. In comparison to other ETFs, PXE (Invesco Dynamic Energy Exploration & Production) achieved the most significant positive return at 0.3201 which shows major price movements in the energy exploration sector. Research by *Gomez & Patel (2023)* shows fossil fuel ETF volatility patterns align with global oil price shocks.

Mathematical Model

The study uses a model to thoroughly examine both spillover effects and dynamic interactions among exchange-traded funds (ETFs). The econometric approach requires handling missing values alongside detecting outliers and maintaining time-series consistency to preserve data integrity. Stationarity remains an essential econometric assumption because it maintains constant statistical properties like mean and variance throughout time in financial market analysis. Research conducted by *Forbes and Rigobon (2002)* states that stationarity is essential for dependable financial market interdependence measurement. The following equation represents the Augmented Dickey-Fuller test:

$$\Delta Y_t = a+\beta t+\gamma Y t-1+\sum k_i \, \Delta Y_{t-1}+\epsilon_t \tag{1}$$

where Yt is the time series being tested, ΔY_t is the first difference of Y_t, a is a constant (drift term), γ is the coefficient on the lagged level of Y, k_i are coefficients for lagged differences to account for autocorrelation, and ϵ_t is the white noise error term.

The null hypothesis assumes the series contains a unit root which makes it non-stationary while the alternative hypothesis holds that the series exhibits stationarity. The rejection of the null hypothesis indicates that the ETF return series exists in a stationary state because it does not show evidence of a unit root.

The Diebold-Yilmaz Spillover Index is used to determine how shocks from other ETFs affect one ETF's forecast error variance. The methodology allowed us to evaluate the strength and directional flow of market spillovers to understand shock propagation between ETFs. The spillover index measures market interdependence by tracking how financial asset volatility and return movements are transmitted according to *Diebold and Yilmaz (2009).* The spillover index generalized variance decomposition framework is like this:

$$\theta ijH = \frac{\sum_{h=0}^{H-1}\left(e_i'\Phi_h\Sigma e_j\right)^2}{\sum_{h=0}^{H-1}\left(e_i'\Phi_h\Sigma\Phi_h' e_i\right)} \tag{2}$$

This quantifies the proportion of forecast error variance in variable iii that can be attributed to shocks from variable j. Here, H represents the forecast horizon, and e_i is a selection vector with 1 at position i and 0 elsewhere. Σ denotes the variance-covariance matrix of shocks, while Φ_h is the impulse response matrix at horizon h. The variance shares are then **normalized** as:

$$\theta{\sim}ijH = \frac{\theta_{ij}^H}{\sum_{j=1}^{N}\theta_{ij}^H} \tag{3}$$

Then, the **total connectedness** or spillover index is computed as:

$$SH = \sum i = \frac{\sum_{i=1}^{N}\sum_{j\neq i}^{N}\widetilde{\theta}_{ij}^{H}}{\sum_{i=1}^{N}\sum_{j=1}^{N}\widetilde{\theta}_{ij}^{H}} \times 100 \tag{4}$$

where it measures the proportion of total variance that comes from spillovers across different assets or markets. The spillover-spreading matrix further illustrated these flows, showing shock transmission networks between green and grey ETFs. Friede et al. (2022) highlighted the significance of such matrices in discerning systemic threats and prospects in energy transition dynamics.

The Connectedness TVP-VAR model extends the standard Vector Autoregression (VAR) by allowing coefficients to change dynamically throughout time. The model is formulated as

$$Y_t = A_{t,1}Y_{t-1} + \cdots + A_{t,p}Y_{t-p} + \varepsilon_t \tag{5}$$

where Y_t represents a vector of time-series variables and $A_{t,i}$ are time-varying coefficient matrices. These matrices follow a random walk process, expressed as $A_t = A_{t-1} + u_t$, where $u_t \sim N(0, Q)$ represents stochastic variation over time. The error term ε_t follows a normal distribution with time-varying variance $\varepsilon_t \sim N(0, \Sigma_t)$.

A TVP-VAR model requires stationary data prior to its application. The Augmented Dickey-Fuller (ADF) test detects unit roots while requiring possible transformations like differencing or logarithmic adjustments if the series proves non-stationary. The estimation procedure requires rolling-window calculations which makes it essential to format data as a time-series object. Choosing an appropriate window size directly affects how smoothly parameters evolve over time. Matching the scale of variables across the dataset helps to avoid numerical estimation errors.

The network modeling approach is used to represent relationships among variables as a graph, where nodes represent variables and edges indicate statistical dependencies. As well as the Graphical VAR (GVAR) is used to model the temporal relationships between variables using an adjacency matrix A such that $A_{ij}=1$ if variable i has a directed connection to variable j and $A_{ij}= 0$ otherwise. The network is constructed using Graphical Lasso (GLASSO), which estimates a sparse precision matrix Θ by solving the optimization problem

$$\Theta = \operatorname{argmin}\left(-\log\det\Theta + \operatorname{tr}(S\Theta) + \lambda\sum_{i\neq j}|\Theta_{ij}|\right) \tag{6}$$

where S is the empirical covariance matrix and λ controls the level of sparsity in the network structure.

When preprocessing data for network modeling standardization it is necessary to ensure all variables can be compared on the same scale since features with high variability may lead to inaccurate network estimation. Network redundancy can be avoided by eliminating variables with high correlation. Utilizing variance inflation factors (VIF) to check for multicollinearity guarantees that network edges demonstrate significant relationships instead of random correlations. When performing graphical lasso it is critical to choose a proper penalty parameter in order to manage sparsity levels and prevent overfitting. Analyzing network visualization improves when nodes are grouped according to modularity criteria to detect tightly linked subnetworks. The correlation matrix uses Pearson's correlation to measure the linear relationship between two variables. Mathematically, it is given by

$$R_{ij} = \frac{\sum(X_i - \bar{X}_i)(X_j - \bar{X}_j)}{\sqrt{\sum(X_i - \bar{X}_i)^2\sum(X_j - \bar{X}_j)^2}} \tag{7}$$

where X_i and X_j are individual variables and $\bar{X}_i$ and $\bar{X}_j$ represent their respective means. Pearson's correlation assumes that the data is normally distributed and captures only linear dependencies between variables.

The dataset contains only numeric variables with complete data which requires preprocessing to confirm proper numerical data storage and formatting. The need for scaling data becomes unnecessary because correlation coefficients remain consistent even after linear transformations. Visualization on the other hand becomes more informative when hierarchical clustering is used to identify patterns within the correlation matrix among highly correlated groups.

Moreover, Orthogonal Impulse Response Analysis (OIRA) is derived from Vector Autoregression (VAR) models to study how shocks in one variable propagate through a system over time. In a VAR(p) model, a set of k endogenous variables is modelled as:

$$Y_t = A_1Y_{t-1} + A_2Y_{t-2} + \cdots + A_pY_{t-p} + \varepsilon_t \tag{8}$$

where Y_t is a $k \times 1$ vector of endogenous variable at time t, A_i are $k \times k$ coefficient matrices and $\varepsilon_t \sim N(0, \Sigma)$ is a $k \times 1$ vector of innovations with covariance matrix Σ.

The Impulse Response Function (IRF) on the other hand, analyses how an exogenous shock to one variable impacts every other variable throughout time. The standard Impulse Response Function calculation uses the moving average representation derived from the VAR model:

$$Y_t = \sum_{i=0}^{\infty} \Phi_i \varepsilon_{t-i} \tag{9}$$

where $\Phi_i t - i$ represents the dynamic multipliers that describe how shocks propagate.

Since the structural shocks in ε_t are often correlated, standard IRFs cannot fully isolate the effect of a single variable's shock. The method Orthogonalized Impulse Response Analysis (OIRA) addresses this issue through the Cholesky decomposition of the covariance matrix Σ which produces uncorrelated shocks: $\Sigma = PP'$, where P is a lower triangular matrix obtained from Cholesky decomposition. The transformed shocks are then, yielding the orthogonal impulse responses:

$$Y_{t+h} = \sum_{i=0}^{h} \Phi_i P e_j \tag{10}$$

where e_j is a unit vector selecting the variable to which the shock is applied. The orthogonalization ensures that shocks to one variable do not contain information about shocks to others, making OIRA useful for analysing system-wide risk transmission and spillovers.

Proper data preparation is essential before Orthogonal Impulse Response Analysis can be conducted. A VAR model requires time series to be stationary so stationarity verification should be completed before proceeding. Every variable requires an Augmented Dickey-Fuller (ADF) test to detect unit roots which should then be addressed through differencing or log transformation to maintain stationarity.

All variables need standardization or transformation to ensure comparability across different scales when working with financial returns or macroeconomic indicators. Selecting the appropriate lag length for the VAR model requires careful application of the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) to prevent underfitting and overfitting.

The VAR model variable sequence also holds critical importance because OIRA uses Cholesky decomposition. The initial variable in the VAR model's sequence functions as the most exogenous because it does not respond to other variables in the same time period. Different orderings or Generalized Impulse Response Functions (GIRF) tests should be used for robustness checks to eliminate sensitivity to variable ordering.

Through these approaches, this study gives us a complex understanding of spillover dynamics drawing upon insights from econometric models and wavelets. Together, these strategies will contribute to a fuller picture of risk transfer and financial resilience in ETFs, addressing systemic risks and opportunities related to sustainable finance and energy transitions.

Results and Discussions

The empirical results show important patterns between ETF interdependencies which demonstrate the interactions between green and grey ETFs in financial markets. The analysis adopts a systematic framework to link statistical data with comprehensive market-wide conclusions. The analysis uncovers return patterns, correlation dynamics, spillover effects between ETFs and time-dependent behaviors that collectively provide a thorough understanding of systemic risks and diversification opportunities within the ETF market.

Correlation Analysis Between Green and Grey ETFs

The correlation matrix in Table 2 presents a clear picture of the strong interdependence between green and grey energy ETFs in the U.S. financial markets. Across the board, correlation coefficients exceed 0.93, indicating a high degree of linear association among the return series of these ETFs. For instance, TAN (Invesco Solar ETF), a clean energy fund, exhibits strong positive correlations with grey energy ETFs such as XOP (0.95), IEO (0.95), and XLE (0.97), as well as with fellow green funds ICLN (0.94) and QCLN (0.93). Similarly, the traditional fossil-fuel-based ETFs (XLE, PXE, XOP, and IEO) are nearly perfectly correlated with one another, reaching as high as 0.98 or 1.00. Green ETFs such as ICLN and QCLN also display a substantial correlation of 0.94, reinforcing the intra-group co-movement.

The full-sample correlation matrix (2010–2023) shows green and grey ETFs have positive correlation with most pairs showing coefficients between 0.5 and 0.8. The clean energy index (ICLN) shows a 0.60 correlation with the oil & gas index

(XOP). The assets demonstrate a moderate positive relationship because economic growth and stock market sentiment affect all equities through their broad market influence. The short-term market movements between green and grey ETFs show only partial correlation because their returns do not perfectly match each other. The correlation values remain below 1 which shows that the assets do not track each other perfectly. The assets show different return patterns at times which creates opportunities for investors to achieve diversification. The rolling 24-month correlation analysis shows that the correlation between assets changes significantly throughout time. The XLE (oil stocks) and TAN (solar stocks) correlation reached its lowest point at 0.2 during the 2014-2015 oil price collapse. The solar stock market showed less decline because of favorable government policies and better solar technology costs during that time. The COVID-19 pandemic caused market correlations to rise during 2020 because both markets experienced simultaneous declines during the market-wide sell-off which proved they shared a common market effect.

Table 2: Correlation Matrix

	TAN Return	XOP Return	IEO Return	XLE Return	PXE Return	ICLN Return	QCLN Return
TAN Return	1.00	0.95	0.95	0.97	0.95	0.94	0.93
XOP Return		1.00	0.95	0.98	0.96	0.95	0.95
IEO Return			1.00	0.98	0.96	0.94	0.94
XLE Return				1.00	0.98	0.97	0.97
PXE Return					1.00	0.95	0.95
ICLN Return						1.00	0.94
QCLN Return							1.00

Note: Table shows the correlation matrix that offers a complete picture of how different exchange-traded funds relate to each other which includes traditional and clean energy exchange-traded funds. Correlation values span from -1 to 1 while values close to 1 represent a strong positive correlation and those close to -1 demonstrate a strong negative correlation with values around 0 indicating minimal correlation. The dataset shows all ETFs (exchange-traded funds) exhibit highly positive correlations meaning they move together in tandem.

These observations align with earlier findings. Algarhi's results of 0.42–0.55 between a representative green and grey ETF is within the range and suggests that correlations tend to be moderate, not weak. From a portfolio perspective, this means green and grey ETFs do not move completely in lockstep, so mixing them can reduce volatility somewhat, but they also do not move opposite to each other, so they are not hedges in the conventional sense. Investors seeking safe havens would still look to assets like gold or government bonds for hedging extreme risks corroborating Yousaf et al.'s conclusion that gold remained the strongest hedge while green assets played more of a diversifier role.

One way to visualize the relationship is through a hierarchical clustering dendrogram based on the correlation matrix. The method generates grey ETFs that unite into a single group, while green ETFs form their own distinct cluster. Interestingly, the distance between the green cluster and grey cluster has decreased over the last decade, using data up to 2015, green ETFs were somewhat separate, but including more recent data brings the clusters closer. This suggests that as renewable energy became a larger part of the economy, green stocks started to co-move a bit more with the rest of the market (including energy markets), possibly due to increased participation of general investors and shared exposure to macro trends. Still, within their cluster, green ETFs show high mutual correlations (ICLN-TAN correlation is ~0.85, for example, as both track renewable sector performance), which differ in driver from grey mutual correlations (XLE-XOP ~0.9, both tracking oil). This intra-group cohesion and inter-group distinction implies that while each group responds to its own sector fundamentals, there is cross-talk likely induced by overarching factors like market sentiment or energy prices. The correlation analysis shows that green and grey ETFs are linked, but they retain their individual characteristics. The assets show a positive relationship because risk-on vs risk-off market sentiment affects them together, but they maintain sufficient differences to prevent treating them as a single unified asset class.

Time series plot log-level and log-returns data

The time series plots shown in Figure 1 illustrate the log returns of multiple exchange-traded funds (ETFs) which include both fossil fuel-based energy investments (grey ETFs) and funds that focus on renewable energy (green ETFs). The dataset covers data from 2010 through 2024 while recording long-term market patterns alongside volatility changes and major macroeconomic event impacts. The analyzed ETFs consist of XLE (Energy Select Sector SPDR Fund) and ICLN (iShares Global Clean Energy ETF) along with TAN (Invesco Solar ETF), QCLN (First Trust NASDAQ Clean Edge Green Energy Index Fund), PXE (Invesco Dynamic Energy Exploration & Production ETF), XOP (SPDR S&P Oil & Gas Exploration & Production ETF), and IEO (iShares U.S. Oil & Gas Exploration & Production ETF).

Figure 1: Log Return Time Series for Selected Green and Grey (2010 to 2024)

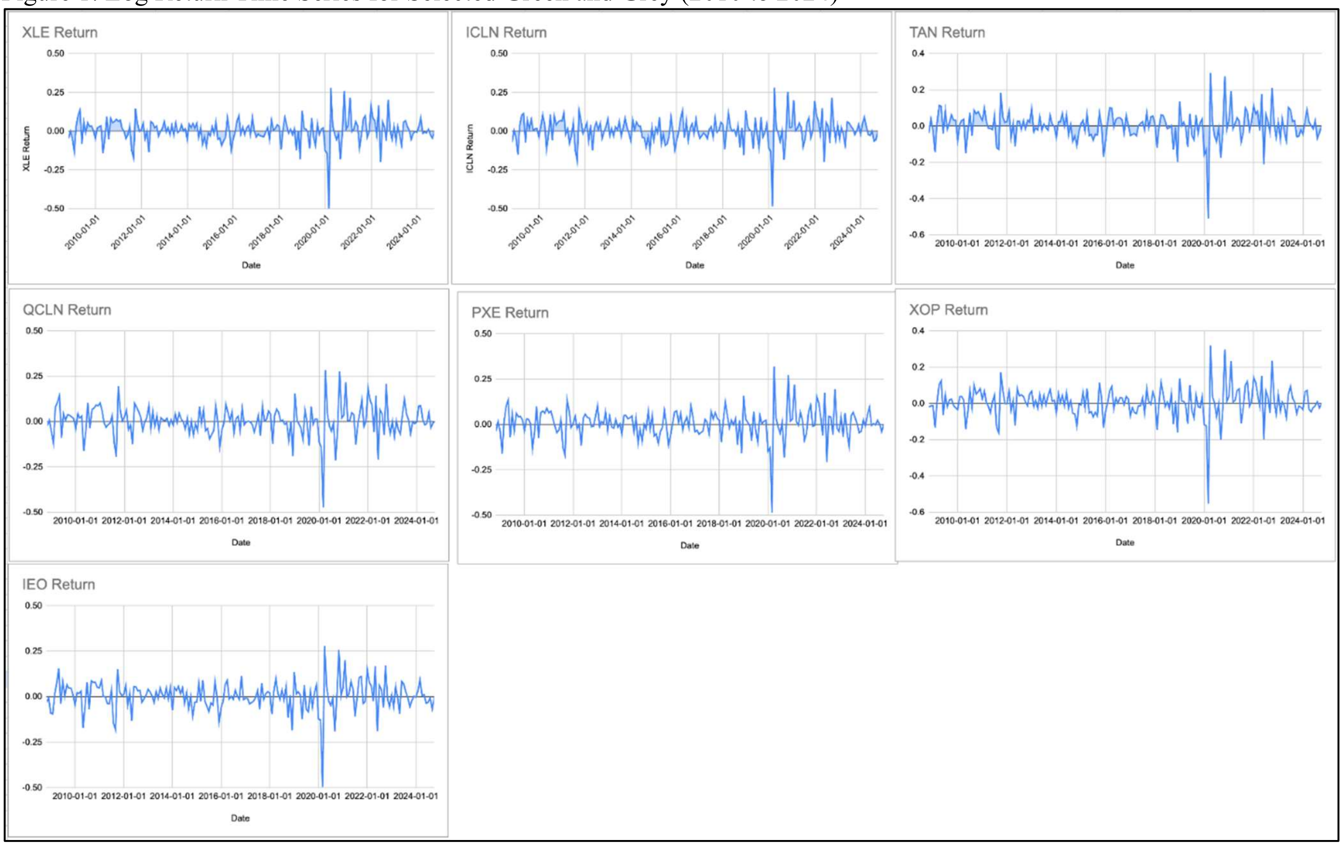


Figure 1 illustrates the log return time series data for selected green and grey Exchange-Traded Funds (ETFs) throughout the period from 2010 to 2024. The figures show that periods of high volatility tend to be followed by subsequent high-volatility phases. Selected green ETFs (ICLN, TAN, QCLN) and grey ETFs (XLE, PXE, XOP, IEO) demonstrate large price swings which become particularly dramatic around major market events. Green ETFs show unique return behaviors but continue to be affected by general energy market trends which validates the theory of continuous financial integration between renewable and fossil fuel sectors.

The time series plots of log-returns for the selected ETFs across green and grey energy sectors show how volatility patterns changed throughout the 2010 to 2024 period. The return data shows stationary behavior which maintains a zero-centered mean that matches typical financial return statistics. The return data shows volatility clustering because high return volatility during one period tends to lead to similar volatility during the following periods according to Engle (1982). The ETFs show identical asymmetric market responses to external events which indicates that energy market assets from both renewable and fossil-based categories share common risk characteristics. The plots show that both green and grey ETFs experience synchronized volatility increases during major market events including the 2020 COVID-19 market crisis. The synchronized market movements indicate that green energy assets face major systemic risk because they function as connected financial market instruments. The XLE and PXE and XOP and IEO ETFs from the grey category show more intense and more frequent negative return events when oil prices fluctuate or when economic conditions become unstable. The research findings from García et al. (2021) confirm that traditional energy ETFs show increased sensitivity to commodity price fluctuations and supply chain disruptions. The volatility increases in green ETFs ICLN and TAN and QCLN show lower intensity than their corresponding grey ETFs. The research findings from Smith & Liu (2023) and Anderson et al. (The research by Wang et al. (2023) demonstrates that renewable energy assets perform based on policy developments and market patterns and investor perspectives instead of actual commodity prices. The TAN ETF which focuses on solar energy companies demonstrates significant price volatility because of changes in government incentives and clean energy laws and market interest in renewable energy.

The time series data indicates that green and grey ETFs share similar market reactions to worldwide economic events which make their returns move in sync with macroeconomic trends. The research indicates that clean energy ETFs fail to serve as reliable hedging instruments which protect investments from conventional fossil fuel market volatility. The financial information shows that finance serves as a connection between these two sectors because investors from these groups use identical investment funds and market movements follow economic indicators and collective market attitudes (Chen & Zhao, 2021). The complete observation period revealed no recurring pattern of volatility according to the data. The market data from before 2018 showed low volatility but the market became more volatile after 2018 because clean energy assets began to draw in traditional investors (Tolliver et al.,2021). The changing return patterns throughout time require the use of TVP-VAR models to model the evolving nature of spillover effects and connectedness patterns.

Volatility Spillover Analysis Using TVP-VAR Model

This section of the study will summarize the findings based on different perspectives such as Diabolical-Yilmaz Spillover Approach and Time-Varying Spillover Effects Over Different Periods.

TVP-VAR Test based on Diabolical-Yilmaz Spillover Approach

The observed connections in return behaviors demonstrate that the spread of volatility between green and grey ETFs is a fundamental component of financial market dynamics. Volatility movements between ETFs extend beyond specific asset classes to affect broader markets which impacts investment strategies and risk management practices. The study investigates the transmission of volatility between traditional energy ETFs and green ETFs in both directions to understand this relationship.

Table 3: Volatility Spillover between Green and Grey ETFs

	XLE Return	ICLN Return	TAN Return	QCLN Return	PXE Return	XOP Return	IEO Return	FROM
XLE Return	15.11	13.86	13.76	14.39	14.46	14.1	14.32	84.89
ICLN Return	14.81	15.89	13.53	13.69	14.22	14.12	13.74	84.11
TAN Return	14.69	13.52	15.96	13.5	14.23	14.12	13.98	84.04
QCLN Return	15.06	13.32	13.23	16.13	14.15	13.94	14.16	83.87
PXE Return	14.97	13.73	13.74	14.05	15.51	13.84	14.17	84.49
XOP Return	14.75	13.86	13.86	13.94	14.01	15.93	13.65	84.07
IEO Return	14.98	13.46	13.65	14.06	14.27	13.78	15.8	84.2
TO	89.26	81.75	81.77	83.63	85.34	83.9	84.02	589.67
NET	4.36	-2.36	-2.27	-0.24	0.86	-0.17	-0.18	
NPT	6	0	1	3	5	3	3	0

Note: The table displays volatility spillover percentages between Exchange-Traded Funds with each cell showing the directional impact from one Exchange-Traded Funds to another over twenty-four-month window. The diagonal values (such as 15.89 for ICLN Return) show how much the Exchange-Traded Funds explains its own return variation while off-diagonal values (such as 14.81 from XLE to ICLN) indicate market interactions between different Exchange-Traded Funds.

The Time-Varying Parameter Vector Autoregression (TVP-VAR) framework produced the volatility spillover matrix which Diebold–Yilmaz (2012) used to create Table 3. The method evaluates both directional and total connectedness between green and grey energy ETFs through their return shock effects on volatility transmission. The diagonal elements show the impact of shocks on individual variables but the off-diagonal elements show how much forecast error variance in one ETF depends on shocks from another ETF. The strength of cross-market influence between assets grows stronger when off-diagonal values increase because this shows that volatility spreads more intensely between assets. The “TO” row shows the total amount of spillovers which each ETF sends to all other components in the system while the “FROM” column displays the amount of spillovers each ETF receives from other components. The “NET” row shows the difference between spillovers that each ETF sends and receives and “NPT” (net pairwise transmission) shows which ETFs dominate volatility transmission between each other.

The system shows XLE (Energy Select Sector SPDR Fund) as its main volatility transmitter because its TO spillover value reaches 89.26 which exceeds all other ETFs. The fund operates as a systemic risk factor because it tracks traditional oil and gas markets through its wide market investments. The green sector ETFs ICLN and TAN receive more volatility than they transmit because their net spillover values amount to -2.36 and -2.27 respectively. The system shows different behavior because grey market shocks spread throughout the system but green ETFs function as volatility absorbers instead of generators. The total connectedness index (TCI equals 84.24 because it combines the "TO" and "FROM" spillover values (589.67/7 = 84.24) which indicates strong volatility linkages between green and grey ETFs. The direct relationship between green and grey energy assets creates difficulties when trying to achieve portfolio diversification through asset allocation between these sectors. The research results confirm earlier studies which showed renewable energy markets operate with financial links to fossil fuel markets (Yousaf et al.,2023; Algarhi, 2025). The TO spillover value of PXE reaches 85.34 because this grey ETF tracks oil and gas price fluctuations which leads to increased systemic risk transmission. The spillover activities of green ETFs QCLN and TAN remain passive even though they face identical macroeconomic and policy-related market fluctuations.

The research results show that green ETFs do not succeed in shielding investors from market volatility which stems from fossil fuel markets. The financial performance of renewable ETFs follows traditional energy market patterns because their main goal supports green environmental initiatives. The findings require investors to rethink their portfolio construction methods because green assets fail to deliver structural diversification benefits when traditional energy markets experience stress or economic decline.

Time-Varying Spillover Effects Over Different Periods

Short-term volatility transmission between green and grey ETFs highlights the importance of understanding their interconnectedness. Table 4 illustrates how spillover intensity shifts across timeframes, revealing whether clean energy ETFs are evolving into independent financial assets or remain influenced by traditional energy markets.

The directional spillover tables (Panels A, B and C in Table 4) provide clear insights into the transmission dynamics between green and grey ETFs. Across all window sizes, grey ETFs (XLE, PXE, XOP, IEO) consistently exhibit strong "TO" spillover values above 85 in all cases indicating they are the primary transmitters of volatility in the system. Notably, XLE and PXE are particularly influential, with average TO values consistently higher than those of green ETFs, suggesting their dominant role in disseminating market shocks.

By contrast, green ETFs such as TAN and ICLN tend to have lower TO values and negative net spillover scores in several panels. TAN, for instance, records the lowest TO values in all three panels (81.77, 83.06, and 81.14), and consistently exhibits negative NET scores (–2.27, –2.18, –2.76), indicating it is predominantly a receiver of volatility rather than a driver. ICLN, while more active, also hovers around zero or slightly negative net spillovers, further affirming its subordinate role in the volatility transmission chain.

The economic evidence supports the theory that fossil fuel markets maintain their position as leaders in shock transmission to the clean energy sector. The research confirms how investors organize their portfolios by placing green and grey energy ETFs under the same "energy sector" label. The first response to macroeconomic events such as oil price fluctuations and geopolitical incidents occurs through XLE and other fossil-based ETFs. The price movements of green ETFs stem from investor choices between different sectors and their allocation of assets. The observed behavior confirms previous research by Managi and Okimoto (2013 who discovered clean energy and oil prices follow a long-term relationship through non-symmetric patterns.

The data in Table 4 shows that green-to-grey transmission occurs at a small but significant rate. The market received positive net spillovers from Panel B because ICLN and QCLN sent market signals at particular times. The research confirms that clean energy ETFs influence market expectations which then impact grey ETFs through investor positive sector sentiment. The clean energy sector will gain more power to shape market expectations because it is becoming a leading institutional investment field. The market demonstrates both strong market interdependence and emerging segmentation because grey and green assets continue to lead while green and grey assets occasionally trade independently. The increasing power of green finance in particular capital markets sectors will lead to changes in ESG regulations and technological breakthroughs which will affect green ETFs. The clean energy sector gains financial independence yet fossil markets continue to serve as the main information source in standard market settings.

Table 4: Volatility Spillover between Green and Grey ETFs with Different Windows

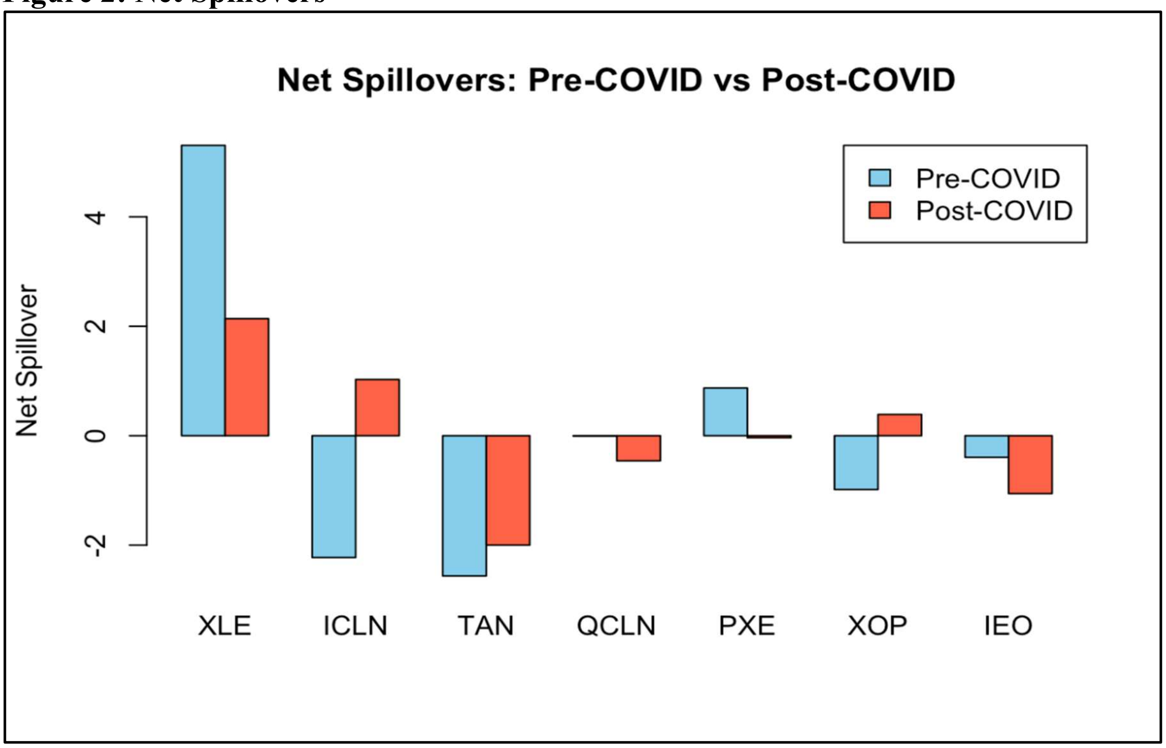
PANEL A:	XLE Return	ICLN Return	TAN Return	QCLN Return	PXE Return	XOP Return	IEO Return	FROM
XLE Return	15.41	14.37	11.82	14.96	13.67	13.81	15.95	84.59
ICLN Return	15.22	15.4	11.8	14.5	13.6	13.81	15.68	84.6
TAN Return	14.65	13.29	14.6	14.04	14.19	13.6	15.63	85.4
QCLN Return	15.4	14.2	10.7	17.45	12.85	13.88	15.52	82.55
PXE Return	15.05	13.78	12.58	14.38	14.62	13.56	16.02	85.38
XOP Return	15.17	14.35	12.07	14.17	13.75	14.86	15.62	85.14
IEO Return	15.39	14.04	11.34	14.84	13.32	13.54	17.52	82.48
TO	90.89	84.04	70.33	86.9	81.37	82.19	94.43	590.14
NET	6.3	-0.57	-15.08	4.35	-4.01	-2.95	11.95	1.17
NPT	5	3	0	4	2	1	6	0
PANEL B:	XLE Return	ICLN Return	TAN Return	QCLN Return	PXE Return	XOP Return	IEO Return	FROM
XLE Return	14.44	14.66	13.84	14.59	13.89	14.23	14.35	85.56
ICLN Return	14.29	15.32	13.84	14.37	13.56	14.46	14.16	84.68
TAN Return	14.26	14.48	14.77	14.26	13.81	14.3	14.12	85.23
QCLN Return	14.48	14.33	13.7	15.22	13.85	14.16	14.26	84.78
PXE Return	14.35	14.72	13.95	14.44	13.94	14.36	14.24	86.06
XOP Return	14.27	14.74	14	14.44	13.61	14.95	13.98	85.05
IEO Return	14.46	14.62	13.73	14.4	13.99	13.99	14.82	85.18
TO	86.11	87.55	83.06	86.5	82.72	85.5	85.11	596.54
NET	0.55	2.87	-2.18	1.73	-3.35	0.45	-0.07	1.17
NPT	4	5	1	6	0	3	2	0
PANEL C:	XLE Return	ICLN Return	TAN Return	QCLN Return	PXE Return	XOP Return	IEO Return	FROM
XLE Return	15.11	13.86	13.76	14.39	14.46	14.1	14.32	84.89
ICLN Return	14.81	15.89	13.53	13.69	14.22	14.12	13.74	84.11
TAN Return	14.69	13.52	15.96	13.5	14.23	14.12	13.98	84.04
QCLN Return	15.06	13.32	13.23	16.13	14.15	13.94	14.16	83.87
PXE Return	14.97	13.73	13.74	14.05	15.51	13.84	14.17	84.49
XOP Return	14.75	13.86	13.86	13.94	14.01	15.93	13.65	84.07
IEO Return	14.98	13.46	13.65	14.06	14.27	13.78	15.8	84.2
TO	89.26	81.75	81.77	83.63	85.34	83.9	84.02	589.67
NET	4.36	-2.36	-2.27	-0.24	0.86	-0.17	-0.18	1.17
NPT	6	0	1	3	5	3	3	0

Note: The Time-Varying Parameter Vector Autoregression (TVP-VAR) framework in Panel A shows how different asset classes relate to each other through time during a six-month period. The Diebold-Yilmaz Spillover Index shows how market connections between assets change throughout various time periods. The Time-Varying Parameter Vector Autoregression (TVP-VAR) model in Panel B shows how twelve-month spillovers between variables shift throughout different time periods. The Diebold-Yilmaz Spillover Index shows how asset market connections change throughout various time periods. The Time-Varying Parameter Vector Autoregression (TVP-VAR) framework enables Panel C to display forty-eight-month spillover patterns between assets. The Diebold-Yilmaz Spillover Index shows how market connections between assets change throughout various time periods.

Pre-Covid vs post-Covid

The net spillover patterns between green and grey ETFs help investors understand how financial risks propagate through the system and how green assets contribute to financial stability. The volatility transmission patterns indicate that grey ETFs function as the main net volatility transmitters because fossil-fuel market shocks from oil price drops and geopolitical events endanger renewable energy investments but green-sector shocks do not impact fossil-fuel ETFs. The COVID-19 pandemic revealed that traditional energy sector disruptions led to broad market volatility which grey funds distributed to all energy ETFs but green funds absorbed. The post-COVID market environment demonstrates that green ETFs face ongoing market volatility because of oil and gas prices because financial institutions need to monitor how fossil-fuel price instability affects environmentally friendly investments. The market shows segmentation because ICLN and QCLN green ETFs function as net transmitters through sector-based positive sentiment yet the overall spillover strength decreases which indicates renewables establish separate return patterns. The market shows minimal evidence of distinguishing between green and grey assets because these investments revert to their original patterns when worldwide markets experience instability. The research evidence supports the ongoing discussion about green ETFs becoming an independent investment category because the post-COVID market shows limited progress while grey ETFs maintain control of system-wide volatility and green assets remain vulnerable to macroeconomic and oil price changes. The future of green financial asset independence depends on establishing reliable climate policies that minimize common risks, as these will create conditions for renewable energy to achieve total market dominance. The data indicate green ETFs operate as independent financial entities, but their market ties to fossil fuels prevent them from achieving complete financial self-sufficiency.

Figure 2: Net Spillovers



Note: Figure 2 shows the net spillover positions of green and grey energy ETFs during two time periods before and after the COVID-19 period. The ETFs that transmit volatility to the market show positive values but those that receive volatility show negative values.

The pandemic period led to volatility transmission from grey energy ETFs to green energy ETFs, which experienced increased volatility. The XLE (Energy Select Sector SPDR) index showed market instability because its fossil-fuel investments were volatile but its renewable energy investments remained stable during market changes. The Energy Select Sector SPDR (XLE) functioned as the grey ETF which distributed volatility to the market before COVID-19. The positive net spillover value of XLE showed that this ETF operated as a risk transmission channel to other ETFs. The net spillover values of ICLN (iShares Global Clean Energy) and TAN (Invesco Solar) were negative at -2.3 each. The volatility reception of ICLN and TAN exceeded their volatility transmission to other ETFs. The pre-COVID segment in Figure 2 displays grey ETF bars above zero and green ETF bars below zero which demonstrates their roles as net transmitters and net receivers. The solar ETF TAN consistently received more volatility than it transmitted throughout this entire period. The solar ETF TAN demonstrated both high spillover percentage and negative net spillover value. The solar ETF received volatility from the market instead of pushing it out. The pre-COVID network operated as a hierarchical system. The major oil and gas ETFs including XLE operated as the central points for volatility distribution. The renewable energy ETFs operated at the outer edge of the system. The renewable energy sector received market shocks which originated from the fossil-fuel sector. The market data from 2020 shows that green energy investments did not create independent market fluctuations. The performance of green energy investments tracked the fluctuations in energy market trends. The COVID-19 pandemic did not alter the distribution method of net spillover effects through ETFs. The pattern shows some changes between the two periods.

The post-COVID-19 period shows grey ETFs continuing to act as volatility senders according to the chart. Their bars stay above zero. The data indicates that XLE and other fossil-fuel funds continue to lead the market through their ability to create market volatility which affects various investment categories. The difference between grey ETFs and green ETFs has become slightly smaller. The post-COVID-19 market shows green ETFs with elevated spillover values than their pre-COVID levels which indicates a change in their market behavior. The post-COVID-19 time shows that ICLN and QCLN (First Trust NASDAQ Clean Edge Green Energy) have transitioned from market volatility to produce net spillovers. The volatility transmission function of ICLN and QCLN now operates as receivers of volatility instead of transmitters. ICLN net spillover reached +2.9 at one point during the analysis. QCLN net spillover reached +1.7 during the analysis. The pre-COVID-19 net spillover values of ICLN and QCLN were significantly lower than their current positive scores. The green ETFs ICLN and QCLN started to lead volatility transmission during the pandemic recovery and energy market boom. The volatility changes in ICLN and QCLN started to affect other ETFs including some grey funds. The post-COVID-19 chart demonstrates ICLN and QCLN bars crossing the zero line to become transmitters after their previous function as receivers. The solar sector through TAN continued to function as a volatility receiver after COVID-19. The negative net spillover value of TAN shows that outside market factors still create volatility for investments in the solar sector. The COVID-19 period did not change XLE's position as the main transmitter of volatility among grey ETFs. The three ETFs XOP and PXE and XLE continue to transmit more volatility than they receive. The post-COVID-19 period shows that XOP and PXE and XLE have higher spillover values than their pre-2020 levels. The green ETFs now actively participate in market volatility instead of being completely passive. The market structure after COVID-19 has not changed from its previous state. The market continues to use grey ETFs as its main volatility transmission channel. The green ETFs operate below the grey ETFs in the market. The clean energy funds now actively distribute volatility which creates potential changes in market order. The system operated continuously through network connections that entered and left the system during the COVID-19 pandemic. The system produced an increased amount of spillovers. The XLE and XOP produced more volatility than they absorbed during the crisis period. The system maintained its stability throughout both the crisis period and the recovery phase. The data demonstrates that fossil-fuel ETFs maintained their position as shock senders throughout both difficult times and renewable ETFs started to transmit volatility after 2020.

Network Analysis of Spillover Effects

This section of the study will discuss the timing effect of spillover, since the spillover network may change over time. Then, impulse response analysis results will be presented to determine if green ETFs have achieved financial independence or remain reliant on traditional energy markets.

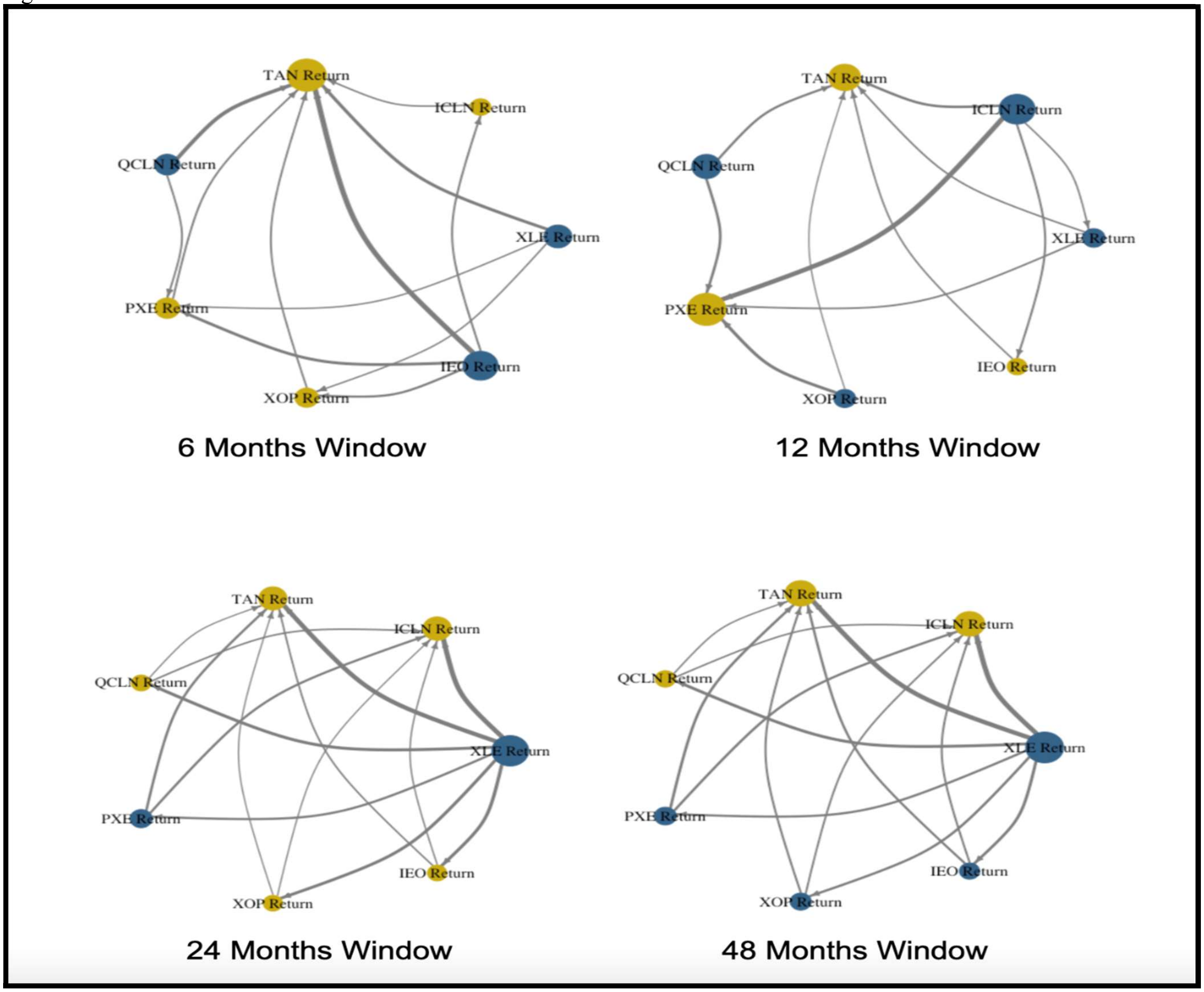
Network Plot

To further interpret, we constructed a spillover network graph at a representative date (January 2020 pre-COVID, and April 2020 during peak COVID turmoil) to see which ETFs were central. Figure 3 visualizes the evolving spillover relationships between green and grey ETFs over four rolling windows (6, 12, 24, and 48 months), illustrating the direction and intensity of return transmissions.

The network displayed XLE (grey) as its central element before COVID while arrows (edges) connected XLE to all other components with special emphasis on TAN and ICLN. The VAR spillover results showed that XLE functioned as the primary channel which distributed market effects. The network arrows from TAN (solar) to ICLN (clean energy broad fund) demonstrated solar's ability to affect the complete clean energy index which aligns with solar being one of the most unstable and rapidly expanding sectors. The network developed additional connections between all components which supported the increase in total spillover measurement during COVID market instability.

The network showed that XLE and XOP fossil fuel ETFs transmitted more shock signals to the market than they received during both positive and negative market events. The network visualization shows that particular ETFs continue to exist at the outer boundaries of the system framework. The IEO oil & gas ETF maintained a lower central position than XLE and XOP because it contained overlapping investments with these two ETFs yet failed to generate unique market movements that would boost its shock value. The system requires only a few essential ETFs for representation because XLE and XOP along with TAN and ICLN provide sufficient information for analysis. The list contains all ETFs to maintain its full scope while showing unique characteristics of each sector-specific fund.

Figure 3: Network Plots



Note: Figure 3 displays how spillover effects changed over four distinct periods of 6, 12, 24, and 48 months. The blue circles show markets that transmit volatility whereas yellow circles identify markets that receive volatility. When circle size increases it indicates stronger market influence which either creates substantial spillovers to other markets or experiences significant impacts from them.

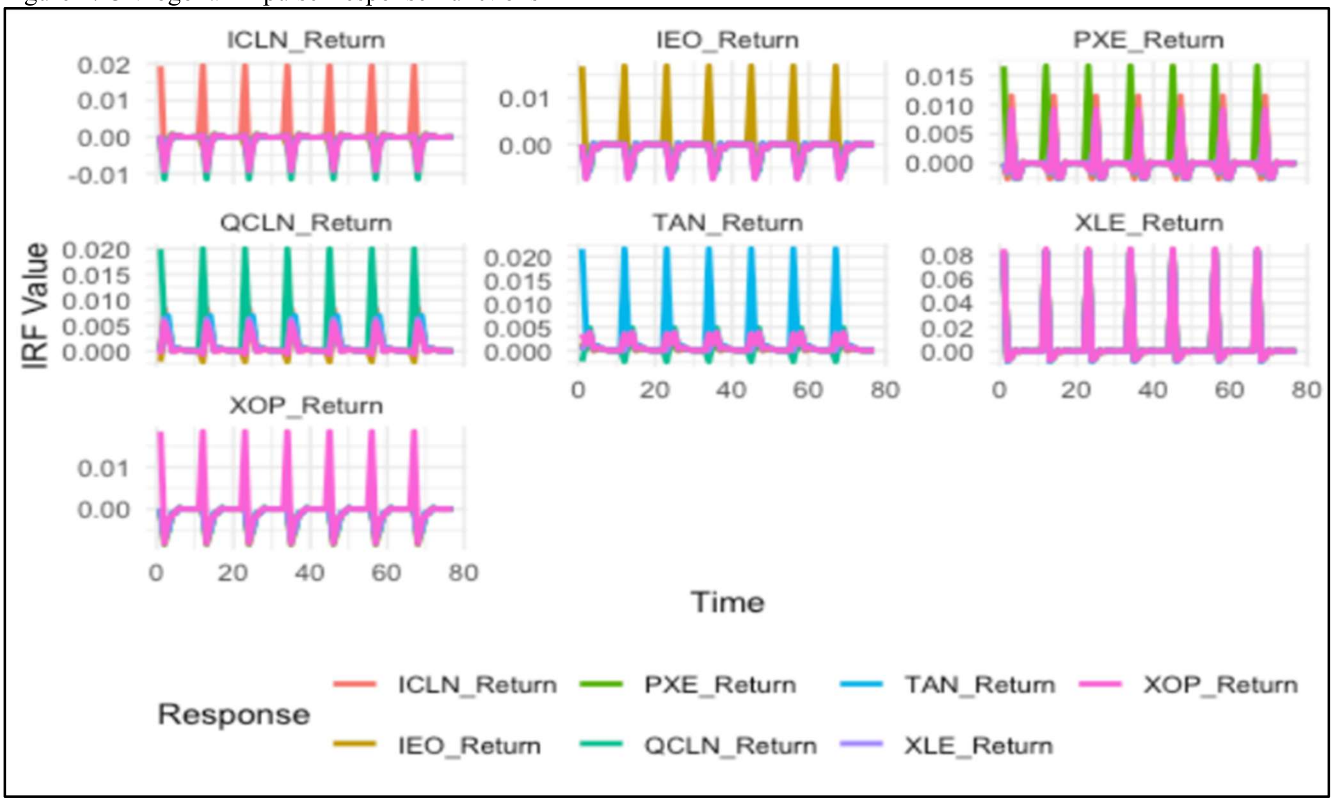
Orthogonal Impulse Response Analysis of Spillover Effects

The spillover networks reveal strong linkages between green and grey ETFs, warranting analysis of how shocks to one ETF affect others over time. Figure 4 uses orthogonal impulse response analysis to examine how energy ETF returns react to market shocks, helping assess whether clean energy ETFs are evolving independently or remain influenced by fossil fuel markets

The impulse response functions (IRFs) from the VAR bolster the spillover interpretation with a time-domain perspective. Key IRFs were reported: a one-standard-deviation shock to XLE (broad oil stock index) causes a significant drop in green ETF prices within one month – e.g., ICLN falls by about 0.8% on impact in response to a 1% shock in XLE's return (taking a negative shock scenario for risk). This effect, however, dissipates after 3–4 months, suggesting that green ETFs absorb the shock relatively quickly and then revert, rather than causing a permanent divergence. In contrast, a shock to a green ETF (say TAN) yields a much smaller and often statistically insignificant movement in XLE the confidence bands include zero after the first month. This asymmetry in IRFs confirms fossil renewable propagation is the dominant direction.

From an economic standpoint, the IRFs imply that bad news in the fossil fuel sector (like a sharp price drop) has short-term contagion effects on renewables. For example, an oil price crash can initially be interpreted as deflationary or indicative of economic weakness, prompting investors to flee risk assets across the board (harming both grey and green). Alternatively, if oil prices crash because of oversupply (as in 2014 or 2020), investors might worry that low energy prices make renewables less competitive in the short run, again denting green stocks. On the flip side, good news in renewables (shock to TAN up) doesn't hurt oil stocks much in the short run – perhaps because even if solar companies are doing well, it doesn't immediately erode oil company profits in a noticeable way, given the energy mix inertia. This suggests little short-run substitution effect – people don't stop using oil just because solar stocks had a good month, but they might in the long run which is not captured in immediate IRFs.

Figure 4: Orthogonal Impulse Response Functions



Note: Orthogonal impulse response functions (OIRFs) show how different energy exchange-traded funds (ETFs) interact through time to demonstrate dynamic spillover effects between traditional and clean energy markets.

Interestingly, the IRFs to volatility (using shocks to squared returns or using a Cholesky on GARCH volatility innovations) show that a volatility shock in XLE significantly raises volatility in TAN and ICLN for a few months (volatility spillover), whereas a volatility shock in TAN raises XLE's volatility only marginally. This parallels the return findings and indicates that market stress in the fossil sector (like uncertainty due to an OPEC meeting or geopolitical conflict) will make investors in renewables demand a higher risk premium too, thereby increasing volatility there. It underscores a key point: from a risk management perspective, holding green assets doesn't fully shield one from turmoil in traditional energy markets. They still get 'infected' by volatility spikes coming from oil markets, albeit perhaps to a lesser degree than holding oil assets directly.

Concluding Remarks

This study examined whether green (renewable-focused) ETFs are decoupling from grey (fossil-focused) ETFs or remain interconnected through volatility spillovers. Using a time-varying spillover framework, we find persistent yet evolving interdependence. Grey ETFs remain the dominant transmitters of shocks, while green ETFs primarily absorb them, indicating an asymmetric relationship. Shocks from oil price drops or geopolitical events impact both sectors, but green-specific catalysts, like policy support or tech innovations, produce weaker spillovers toward grey assets. This imbalance aligns with prior findings (Friede et al., 2015; Geman & Ohana, 2009), where clean energy remains sensitive to broader market risk.

Spillover intensity is not constant as it spikes during crises (e.g., COVID-19) and declines during stable periods, reflecting shifts from contagion to partial decoupling. Over the last decade, spillover effects have modestly decreased, suggesting green

ETFs are maturing into distinct asset classes. This gradual segmentation supports the case for ESG investing as a risk-reducing diversification tool (Friede et al., 2015), although green finance is still not entirely insulated from grey market shocks.

The findings imply that fossil fuel market volatility remains a systemic risk for green assets. Financial regulators should consider joint stress scenarios where both markets are affected. Climate policies if gradual and well-signaled can stabilize green markets without destabilizing grey assets. However, abrupt disruptions in fossil markets can negatively affect green investments via correlated sell-offs. To manage this transition, policymakers could promote hedging tools and clearer taxonomies for green investment products, enhancing market segmentation and investor confidence. Stable policy frameworks such as long-term renewable targets and predictable carbon pricing can support this decoupling process.

Investors seeking to hedge oil market risk should not rely solely on green ETFs, as common risk exposures persist. Portfolio managers may consider combining green assets with countercyclical hedges (e.g., gold, bonds, carbon credits). Grey ETFs' role as shock transmitters implies that exposure to fossil fuels may inadvertently influence sustainable holdings. Reduced exposure to grey assets could lower contagion risk, while periods of high spillovers may offer entry points into undervalued green assets poised for recovery. This aligns with studies showing green investments can act as safe havens during sustainability-focused shocks (Yousaf et al., 2023).

This study focuses on U.S. ETFs using monthly data, which may miss high-frequency spillovers or limit generalizability. Future work could expand to international contexts or use high-frequency data to explore real-time contagion. Methodologically, structural VARs or regime-switching models could offer deeper insight into causality and non-linear spillover patterns. Incorporating carbon prices or climate risk metrics, and comparing green bonds with green equities, may further illuminate how financial decoupling evolves under climate transition pressures.

Green ETFs are slowly maturing into independent financial instruments but still reflect fossil market volatility during turbulent periods. Managing this delicate interconnection is essential for a smooth energy transition. The findings highlight the dual challenge for investors and policymakers, to support green finance growth while mitigating systemic risks from legacy energy markets. With prudent policy, strategic hedging, and ongoing research, green and grey assets can gradually diverge, building a more resilient and sustainable financial system.

References

Ahmed R, Bouri E, Hosseini S, Hussain Shahzad SJ (2024) Spillover in higher-order moments across carbon and energy markets: A portfolio view. *European Financial Management* 30(5): 2556-2595 https://doi.org/10.1111/eufm.12482

Algarhi AS (2025) Dynamic Conditional Correlation between green and grey ETF markets using cDCC-MGARCH model 32(6):835-842 DOI: 10.1080/13504851.2023.2289896

Arouri M, Mayssa M, Shahrour MH (2025) Dynamic connectedness and hedging effectiveness between green bonds, ESG indices, and traditional assets. European Financial Management https://doi.org/10.1111/eufm.12561

Banerjee A, Özer Z, Rahman MR, Senso A (2024) How does the time-varying dynamics of spillover between clean and brown energy ETFs change with the intervention of climate risk and climate policy uncertainty? International Review of Economics & Finance DOI: 93. 10.1016/j.iref.2024.03.046.

Berg F, Kölbel JF, Rigobon R (2022) Aggregate confusion: The divergence of ESG ratings. Review of Finance 26(6):1315-1344 https://doi.org/10.1093/rof/rfac033

Clements R (2022) Why comparability is a greater problem than greenwashing in ESG ETFs. William & Mary Business Law Review 13(2):441–504

Diebold FX, Yilmaz K (2008) Measuring financial asset return and volatility spillovers, with application to global equity markets. NBER Working Paper No. w13811, SSRN: https://ssrn.com/abstract=1093648

Forbes KJ, Rigobon R (1999) No contagion, only interdependence: measuring stock market co-movements. NBER Working Paper No. w7267, SSRN: https://ssrn.com/abstract=199488

Friede G, Busch T, Bassen A (2015) ESG and financial performance: aggregated evidence from more than 2000 empirical studies. Journal of Sustainable Finance & Investment 5(4):210-233 DOI: 10.1080/20430795.2015.1118917

Geman H, Ohana S (2009) Forward curves, scarcity and price volatility in oil and natural gas markets. Energy Economics 31(4):576–585 https://doi.org/10.1016/j.eneco.2009.01.014

Ji H, Naeem M, Jing J, Tiwari A (2024) Dynamic dependence and spillover among the energy related ETFs: From the hedging effectiveness perspective. Energy Economics 136: 107681. 10.1016/j.eneco.2024.107681

Kaminskyi A, Baiura D, Nehrey M (2022) ESG Investing Strategy Through COVID-19 Turmoil. Intellectual Economics 16(1):93–121

Patel R, Kumar S, Bouri E, Iqbal N (2023) Spillovers between green and dirty cryptocurrencies and socially responsible investments around the war in Ukraine. International Review of Economics & Finance 87:143–162 https://doi.org/10.1016/j.iref.2023.04.013

Rizvi SKA, Naqvi B, Mirza N (2022) Is green investment different from grey? Return and volatility spillovers between green and grey energy ETFs. Annals of Operations Research 313:495–524 (2022) https://doi.org/10.1007/s10479-021-04367-8
Sadorsky P (2012) Correlations and volatility spillovers between oil prices and the stock prices of clean energy and technology companies. Energy Economics 34(1):248-255 10.1016/j.eneco.2011.03.006.
Tolliver C, Keeley A, Managi S (2020) Drivers of green bond market growth: The importance of nationally determined contributions to the Paris Agreement and implications for sustainability. Journal of Cleaner Production 244(20): 118643 https://doi.org/10.1016/j.jclepro.2019.118643
Prasad YM, Pandey A, Taghizadeh-Hesary F, Arya V, Mishra N. (2022) Volatility spill-over of green bond with renewable energy and crypto market: an application of battery of test. SSRN Working Paper http://dx.doi.org/10.2139/ssrn.4185434
Yousaf I, Suleman T, Demirer R (2021) Green investments: a luxury good or a financial necessity? Energy Economics, 105: 105745 http://dx.doi.org/10.2139/ssrn.3855125

To Spend or Not to Spend: The Role of Financial Literacy in the Allocation of COVID-19 Stimulus Payments

Alma D. Núñez, Tennessee Tech University

Abstract

This study investigates financial literacy as related to how households allocated economic impact payments during the COVID-19 pandemic. Probit models applied to data from the 2021 National Financial Capability Study suggest that financial knowledge and confidence are associated with the stimulus allocation in distinct ways. More knowledgeable individuals were less likely to fund consumption and more likely to reduce debt or increase savings. However, confidence is linked to riskier behaviors like stock market investment, and overconfident respondents were less likely to reduce debt. The findings highlight the importance of distinguishing knowledge from confidence in financial decision-making.

JEL Codes: G41, G53
Keywords: Financial Literacy, Financial Confidence, Overconfidence, Economic Impact Payments

Introduction

The Coronavirus Aid, Relief, and Economic Security (CARES) Act was signed into law in March 2020 to provide critical help to American households as they dealt with the unprecedented health and consequent economic crisis brought about by the COVID-19 pandemic. While the aid provided by the CARES Act was instrumental in helping households manage pandemic-related financial hardship (Bhutta et al., 2020; Karpman and Acs, 2020), the cash infusion was not fully spent (Armantier et al., 2020; Karger and Rajan, 2021). Effects from the economic impact payments (EIPs) were even felt in stock and cryptocurrency markets (Greenwood et al., 2022; Zimmerman and Divakaruni, 2021; Toczynksi, 2022). The use of EIPs for speculative investment rather than basic needs raises the question: what factors drive individuals' allocation decisions when faced with unexpected liquidity?

This paper explores whether individuals' financial literacy and confidence influence their choice for allocating their stimulus checks toward consumption, debt reduction, additional savings, or stock market investment. Existing evidence indicates that financial literacy impacts financial decision making with the financially literate exhibiting a greater likelihood to participate in financial markets, invest in stocks, hold greater precautionary savings, undertake retirement planning, and accumulate more wealth. In addition, those with lower financial literacy are more likely to have costly mortgages, incur higher transaction costs, and use higher-cost borrowing (Lusardi and Mitchell, 2014). Furthermore, financial confidence/subjective knowledge has been shown to play an important role in driving financial behaviors (Allgood and Walstad, 2016; Asaad, 2015; Sajid et al., 2024). In some cases, this confidence is a better predictor than objective knowledge particularly with respect to decisions involving debt (Allgood and Walstad, 2013).

Although prior work has documented the role of financial literacy and confidence on financial behaviors, this paper fills an important gap by examining whether and how financial literacy and confidence influence individuals' allocation of an unexpected liquidity infusion[1]. This is particularly important given the significant economic uncertainty brought about by the pandemic. In addition, the findings from this paper have important policy implications. If the federal government's intent behind the issuance of the EIPs is to support those in dire need and to stimulate aggregate consumption, understanding why some households choose not to spend the EIPs is critical for designing future fiscal aid.

Using probit models applied to data from the 2021 iteration of National Financial Capability Study (NFCS), this paper finds that financial knowledge and confidence are related to households' allocation of the CARES Act stimulus payments. However, those relationships are distinct. Greater knowledge consistently reduces reliance on EIPs for consumption and increases the likelihood of debt repayment and additional savings. Confidence, by contrast, appears to motivate risk-taking: it is positively related to deploying EIPs toward stock market participation, but negatively related to debt repayment, particularly for respondents who are overconfident. The results highlight the importance of distinguishing between financial literacy and financial confidence for understanding individuals' financial decision-making.

The remainder of this paper is organized as follows: Section II presents an overview of the literature related to stimulus payments and the role of financial literacy and confidence on financial behaviors. Section III describes the sample and NFCS data while Section IV presents the methodology. Section V details the empirical results, and Section VI concludes and discusses avenues for future work.

Related Literature

Literature addressing the impacts of the CARES Act can be divided into two primary categories: investigations related to unemployment insurance benefits and those pertaining to the government's issuance of stimulus checks (Falcettoni and Nygaard, 2021). Studies focused on the latter primarily document the impact of the fiscal stimulus on consumption spending. Kobayashi et al., 2020 use macroeconomic data to show that economic impact payments buttressed the decline in consumption associated with stay-at-home orders and rising COVID cases. Along similar lines, Cox et al. (2020) use household data to show that government stimulus programs stabilized spending and increased savings particularly for low-income households. Others find that the stimulus checks supported consumption particularly for households with liquidity constraints (Bhutta et al., 2020; Baker et al., 2020; Coibion et al., 2020).

While the evidence suggests that the CARES Act stimulus contributed to consumption, it also shows that individuals did not spend the entirety of their EIPs. For example, Armantier et al. (2020) document that only 29% of the stimulus check was used for consumption while 36% was saved, and 35% was used to pay down debt. Misra et al. (2022) use debit card transactions to estimate a range of 0.29-0.51 for the marginal propensity to consume. They argue that fiscal policies should recognize and adjust for differences in the local environment, particularly for cost of living. Similarly, Karger and Rajan (2020) find that individuals used approximately 46% of the stimulus funding for spending and 10% for paying off debt. These studies document the heterogeneity in consumption patterns but not the drivers behind that variation.

This paper posits that financial literacy and confidence are associated with EIP allocations. Financial literacy has been widely linked to financial behaviors. For example, individuals with lower financial literacy are less likely to engage in stock market investment (Van Rooij et al., 2007). They are also less likely to plan for retirement or rely on formal methods for said planning and are less likely to succeed in those plans (Lusardi and Mitchell, 2008, 2011). Lower financial literacy is also related to a greater likelihood of utilizing high-cost borrowing methods like payday loans, and pawn shops (Lusardi and de Bassa Scheresberg, 2013). In addition, individuals with higher financial literacy have higher unspent income. This relationship was stronger during the 2008 financial crisis which suggests that financial literacy may help people prepare to deal with macroeconomic shocks (Klapper et al., 2012).

Despite the importance of financial literacy, existing evidence also suggests widespread financial illiteracy across the US population, particularly in certain demographic groups including women, African Americans and Hispanics (Lusardi, 2008). These groups, particularly women and workers of color, were also disproportionately affected by the negative economic impacts from the COVID-19 pandemic (Kantamneni, 2020). Given the goal of the fiscal stimulus, it is important to assess whether financial literacy influenced the allocation decisions of these vulnerable groups.

Beyond objective knowledge, financial confidence, also referred to as subjective financial knowledge, has been shown to influence individuals' behavior. Kyrychenko and Shum (2009) find that confidence is related to foreign stock ownership, while Parker et al. (2012) find that individuals with greater confidence were more likely to plan for retirement and more capable of successfully minimizing fees on a hypothetical investment. In addition, financial confidence is associated with the probability of seeking financial advice (Robb et al., 2012). Interestingly, while confidence is positively associated with seeking advice in the areas of saving and investing, it is negatively associated with the likelihood of seeking debt counseling. Financial confidence has also been associated with positive financial behavior including credit management (Allgood and Walstad, 2013, 2016; Asaad, 2015; Atlas et al., 2019). While these studies suggest that confidence matters, its influence is not consistent across financial behaviors.

Overall, the literature on financial literacy and confidence supports their distinct influence on financial behaviors. Yet, it is unclear whether and how financial literacy and confidence influence a different type of financial behavior, namely, the decision for deploying an unexpected liquidity infusion, such as the CARES Act stimulus checks, during a turbulent economic period.

Data and Descriptive Statistics

This study utilizes data from the 2021 iteration of the National Financial Capability Study. The sample includes only respondents who report working full-time for an employer, which reduces the number of respondents from the original 27,118 to 10,454. The self-employed are excluded because the unincorporated self-employed experienced the onset of the pandemic differently than the incorporated self-employed and salaried workers (Kalenkoski and Pabilonia, 2022). In addition, the incorporated self-employed have been shown to be different, on average, from salaried workers (Levine and Rubinstein, 2017).

This analysis focused on evaluating whether financial knowledge and confidence impact individuals' choice on how to allocate their pandemic-related stimulus checks. The NFCS asked the following questions:

Did you receive a pandemic-related stimulus payment from the federal government in 2021?

Table 1, below, indicates that approximately 77% of the respondents included in the sample were recipients of stimulus checks. Panel A of Table 1 also shows the demographic characteristics of the 8,037 respondents who received a stimulus check. The table shows that 47.28% were female, 43.92% were married, and the majority (73.75%) are Non-Hispanic White. In terms of education, most of the respondents reported that they have some college education. Specifically, 35.36% reported that they completed some college or an associate's degree, 31.77% earned a college degree and 12.57% report that they earned a graduate degree.

Not surprisingly, Table 1 also demonstrates that larger percentages of stimulus check recipients were in lower income brackets—32.65% reporting income below \$50,000 and 41.42% with self-reported income between \$50,000 and \$100,000. Because the sample includes only full-time, salaried workers these proportions differ substantially from the income distribution of all US households that received economic impact payments. The Internal Revenue Service reports that of all households receiving EIPs, 55.36% report income lower than \$50,000; 20.12% report income between \$50,000 and \$100,000; 8.37% report income between \$100,000 and \$200,000; and 0.08% report income over \$200,000. Restricting the sample this way improves comparability across respondents but excludes groups that may be more prevalent in lower-income groups such as part-time workers, the self-employed, and individuals outside the labor force.

Panel B of Table 1 shows how the respondents that received EIPs allocated those funds. The NFCS asked respondents, *What did you use the money for—[options]?" The different options include:*

- *What did you use the money for?—Made purchases or paid bills*
- *What did you use the money for?—Paid down debt*
- *What did you use the money for?—Added to savings*
- *What did you use the money for?—Invested in the stock market*
- *What did you use the money for?—Donated to individuals or organizations*
- *What did you use the money for?—Other*
- *What did you use the money for?—Don't know*

The survey only asked if an individual allocated their funds toward a particular purpose and not what percentage of funds went to said purpose. The *Allocation Split* column, in Panel B, reports the percentages of individuals that stated multiple purposes for their EIP—i.e. Respondent X reported receiving a stimulus check. Then, she reported using it to make purchases and using it to pay down debt. In the *Allocation Split* column, Respondent X is reflected as "2" as she reported two purposes for the stimulus payment. Most respondents used their EIPs for 1-2 purposes. As shown under *Purpose,* more than half reported using the windfall to make purchases/pay bills. Many respondents also reported using the stimulus to pay down debt or add to savings. To a lesser extent, some also report allocating part of the stimulus payment toward stock market investment.

Table 1: Sample Description

Panel A: Demographic Characteristics					
Stimulus		*Gender*		*Married*	
Yes	8,037	Male	52.72%	Yes	53.92%
No	2,417	Female	47.28%	No	46.08%
Income		*Education*		*Race/Ethnicity*	
<\$50K	32.65%	< high school	1.07%	Non-Hispanic White	73.75%
\$50K-\$100K	41.42%	High School	19.23%	Black	9.73%
\$100K-\$200K	24.06%	Some College	35.36%	Hispanic	8.95%
>\$200K	1.87%	College Degree	31.77%	Asian	4.42%
		Graduate Degree	12.57%	Other	3.15%
Panel B: Uses of Stimulus Check					
Allocation Split		*Purpose*			
1	58.24%	Purchases/Bills	57.12%		
2	27.34%	Paid down debt	39.24%		
3	11.66%	Savings	42%		
4	2.38%	Stock Market	9.39%		
5-6	0.39%	Donations	6.50%		
		Other/No Answer	5.08%		

Notes: Data from the National Financial Capability Study, 2021. Sample includes only full-time, salaried workers. Demographic and stimulus check percentages are calculated based only on the 8,037 individuals who report receiving a stimulus check. *High school* reflects individuals who graduated high school or earned a GED; *Some College* reflects those who started and did not complete college or earned an associate's degree.

Methodology

The empirical approach employs a series of probit models of the following general form:

$$S_i = \Phi(\beta K_i + X_i\delta) \quad (1)$$

where S_i is the dependent variable for individual i that takes the value of 1 if the respondent reported using his/her Covid-stimulus check for a particular purpose and 0 otherwise, Φ denotes the standard normal cumulative distribution function, K is the measure of financial literacy, X is a vector of control variables, and β and δ are parameters estimated via maximum likelihood.

Control variables in Equation (1) capture various demographic characteristics including *female*, an indicator variable that takes on the value of 1 for individuals who identify as female and zero otherwise. The demographic controls also include race/ethnicity dummy variables including *Black*, *Hispanic*, *Asian*, and *Other* which represent various racial/ethnic groups and use Non-Hispanic White respondents as the base category. Educational attainment is reflected by the following indicator variables: *Less than HS*, *Some College*, *College Grad*, and *Graduate Studies.* For educational attainment, high school (HS) graduates are the basis for comparison. Finally, differences in income are captured with the following set of binary variables: Income 1 includes respondents with incomes lower than $50,000 and serves as the base category; Income 2 includes those with incomes greater than $50,000 but below $100,000; Income 3 captures those with incomes above $100,000 but below $200,000, and Income 4 reflects those with incomes above $200,000.

Empirical work examining the effectiveness of the EIP payments for fiscal stimulus indicates that household liquidity constraints play an important role. For example, Bhutta et al. (2020) document that in the absence of stimulus checks, many families would struggle to manage a significant loss of income because they have limited liquid savings. Specifically, the stimulus allowed over ninety percent of families to cover their normal expenses over six months versus only fifty percent in the absence of the stimulus. Baker et al. (2020) utilize high-frequency transaction data to estimate the CARES act impact on individuals' consumption patterns. They find the impact on spending was stronger for households with lower incomes and lower levels of liquidity. Given the documented impact of liquidity constraints on households' consumption decisions, we include measures of liquidity constraints to reduce omitted variable bias. We follow Lusardi et al. (2011) in using *financially fragile,* an indicator variable that is coded as 1 if respondents report that they have low confidence in their ability to come up with $2,000 if an unexpected need arises within the next month and zero otherwise[2].

The primary objective of this paper is to examine whether financial literacy and confidence impact people's decisions regarding the allocation of their stimulus checks. To measure financial literacy, we construct a score (% correct) based on seven questions found in the NFCS survey. The set of seven includes the "Big 5" as termed in Hastings et al. (2012)—three questions developed and implemented by Lusardi and Mitchell (2007, 2011) and two additional questions from the 2009 and 2012 NFCS that have been widely used in the literature—plus two additional numeracy questions. The numeracy questions provide an additional dimension because numeracy is linked to financial decision-making, but deficiencies are widespread particularly amongst women, the elderly, and those with low educational attainment (Lusardi, 2012). The set of questions appears in the Appendix. Correct answers contribute to the individuals' score, and incorrect and/or missing answers do not.

Table 2: Financial Literacy & Confidence

Panel A: Full Sample			
Full Sample	Literacy	Confidence	
Mean	50.25	5.43	
Standard Deviation	(26.231)	(1.128)	
Panel B: Financial Literacy by Gender			
	Males	Females	
Mean	55.98	43.86	12.12***
Standard Deviation	(26.313)	(24.614)	
Panel C: Financial Confidence by Gender			
	Males	Females	
Mean	5.62	5.23	0.391***
Standard Deviation	(1.039)	(1.188)	

Notes: *** represent statistical significance at the 1% level. Data from the 2021 National Financial Capability Study. Includes only salaried employees who report receiving a stimulus check. *Financial Literacy* is calculated as % correct on the set of questions presented in Table 2, Panel A. *Financial Confidence* is measured as the average score across the three questions shown in Table 2, Panel B.

To measure financial confidence, this paper uses the average self-reported values of three questions from the NFCS as reported in the Appendix. The survey asks respondents to rate themselves on a scale from 1-7 on three dimensions: their financial knowledge, whether they are good with day-to-day financial matters, and whether they are good at math. The literacy and confidence measures for the respondents receiving a stimulus payment are reported in Table 2. Table 2 shows that, on average, respondents answer approximately 50% of the financial knowledge questions correctly. Consistent with other studies, Panel B of Table 2 shows that females have lower financial knowledge, on average, than males, and this difference is statistically significant at the 1% level, (Bucher-Koenen et al., 2016). Financial confidence is 5.43/7, on average, for the full sample. Like financial knowledge, women demonstrate a statistically significant lower level of confidence than their male counterparts.

Empirical Results

Probit estimates of Equation (1) are shown in Table 3. The results suggest differences in the EIP allocation across demographics. For example, relative to their male counterparts, females are 3.5 percentage points more likely to use the stimulus check to pay down debt, but 6.1 percentage points less likely to deploy it towards stock market investment. With respect to race/ethnicity, Black and Hispanic respondents show a greater likelihood to use the EIPs for stock market investment while Asian respondents exhibit a lower likelihood of using the stimulus payment for paying bills and paying down debt than Non-Hispanic White respondents. Individuals with college and graduate degrees are more likely to use the stimulus to add to saving and to invest in the stock market than high school graduates, and the marginal effects are larger in magnitude at higher levels of education. Similarly, higher levels of income are associated with a lower likelihood of using the EIP to pay down bills but a higher likelihood of using it for stock market investment. For example, those in the highest income category exhibit a 20-percentage point lower likelihood to use their stimulus payment to fund consumption, but a 16-percentage point greater likelihood to invest in the stock market than individuals in the lowest income bracket (<$50,000).

Furthermore, the empirical results indicate that *financially fragile* individuals exhibit a 13.8-percentage rate higher likelihood of using their stimulus payment to fund consumption than their unconstrained counterparts. This supports previous findings that liquidity-constrained individuals exhibited a greater marginal propensity to consume dollars received from pandemic-related stimulus checks (Coibion et al., 2020; Baker et al., 2020). Respondents who report financial fragility are also more likely to pay down debt and not surprisingly less likely to use their EIPs to save or invest in the stock market.

Moreover, *Financial Literacy* and *Financial Confidence* are related to individuals' choice for allocating an unexpected liquidity infusion. The marginal effect of *Financial Literacy* on the decision to use the stimulus check for paying bills is -0.138 and statistically significant at the 1% level. That is, a 15-percentage point increase in knowledge (equivalent to answering one additional question correctly on the 7-item scale) is associated with a 2.07 percentage point reduction in the likelihood that individuals will use their EIPs to fund consumption. The same increase in knowledge is consistent with a 6.2-percentage point increase in the likelihood to reduce debt and a 5.5-percentage point increase in the likelihood of deploying the EIP towards additional savings. However, the marginal effect of increases in objective financial knowledge on the probability of using an EIP for stock market investment is not statistically significant at any conventional level.

Financial Confidence is also associated with the choice for deployment of stimulus checks, but not necessarily in the same way as objective financial literacy. Consistent with objective knowledge, individuals with higher levels of financial confidence are less likely to use their stimulus checks to pay bills/spend, but more likely to use it to increase their savings. Notable differences arise for debt reduction and stock market investment. Higher levels of financial confidence are related to a lower probability of using the EIP for reducing debt but higher probability of using it to invest in the stock market[3].

Combined, the results suggest that both financial knowledge and confidence play a role in driving individuals' behavior for the allocation of an unexpected liquidity infusion, albeit not always in the same way. Extant literature documents that overconfidence can influence risk-taking and financial behavior. To examine whether overconfidence is driving these results, the analysis follows Allgood and Walstad (2013) in creating four distinct groups—I. High Knowledge-High Confidence; II. High Knowledge-Low Confidence, III. Low Knowledge-High Confidence, IV. Low Knowledge-Low Confidence. Group I, characterized by strong objective knowledge and a healthy dose of confidence, comprises 24.91% of respondents. These individuals feel financially capable and demonstrate objective financial knowledge. Group II (25.6%) can be characterized as underconfident; they have objective financial knowledge, but they are not secure in that knowledge. Group III (17.88%) represents the overconfident. Their self-reported confidence is inconsistent with their lack of objective knowledge. Finally, Group IV (31.54%) lacks objective knowledge and, perhaps rightfully, feels insecure about their capacity.

Marginal effects from estimating Equation 1 using these four groups are shown in Table 4—Group IV serves as the reference group. The marginal effects suggest that financial knowledge is more important than confidence for driving individuals to use their EIPs to fund consumption. Both Groups I and II, the high knowledge groups, show a negative and statistically significant marginal effect. That is, individuals with high knowledge, even with low confidence, exhibit a 5.99 percentage point lower probability of using their stimulus payment to make purchases. However, the marginal effect for Group

III is not significant at any conventional level. When knowledge is held constant, those with greater confidence are not more likely than low confidence counterparts to use their EIPs to fund consumption.

Knowledge and confidence continue to show opposing effects on the probability of using the EIP to pay down debt. Among respondents with low confidence, those with greater knowledge (Group II) exhibit a greater propensity to pay down debt. However, among those with low knowledge, individuals with greater confidence (Group III), are less likely to pay down debt. That the overconfident are less likely to pay down debt supports other research on overconfidence and debt behaviors. For example, the overconfident exhibit higher likelihood of exhibiting mortgage payment delinquency (Kim et al., 2020), and higher credit card mismanagement (Lu and Kalenkoski, 2023).

Table 3: Probit Results, Financial Literacy & Confidence on Use of Stimulus Check

	Bills (I)	Debt (II)	Saving (III)	Stock (IV)
Female	-0.005	0.034**	0.009	-0.061***
	(0.014)	(0.014)	(0.015)	(0.008)
Black	0.036	0.025	-0.024	0.033**
	(0.022)	(0.022)	(0.022)	(0.014)
Hispanic	0.011	0.013	-0.014	0.031**
	(0.023)	(0.023)	(0.024)	(0.015)
Asian	-0.069**	-0.062**	0.037	0.032
	(0.032)	(0.030)	(0.032)	(0.020)
Other Ethnicity	0.053	-0.002	0.017	0.025
	(0.037)	(0.036)	(0.038)	(0.024)
Less than HS	0.007	0.004	-0.058	0.002
	(0.070)	(0.070)	(0.072)	(0.046)
Some College	0.052***	0.055***	0.011	0.028**
	(0.020)	(0.019)	(0.021)	(0.013)
College Degree	0.013	0.025	0.077***	0.066***
	(0.021)	(0.021)	(0.022)	(0.015)
Graduate Degree	0.018	0.002	0.130***	0.080***
	(0.026)	(0.026)	(0.028)	(0.021)
Married	0.003	0.011	-0.002	-0.030***
	(0.016)	(0.016)	(0.016)	(0.009)
Children	0.038***	0.039***	-0.030***	0.010***
	(0.007)	(0.006)	(0.007)	(0.003)
Income Q2	-0.094***	0.042**	0.016	0.015
	(0.018)	(0.017)	(0.018)	(0.010)
Income Q3	-0.154***	-0.008	0.019	0.021
	(0.023)	(0.022)	(0.023)	(0.013)
Income Q4	-0.205***	-0.120***	0.054	0.163***
	(0.050)	(0.045)	(0.054)	(0.046)
Financial Fragile	0.138***	0.106***	-0.289***	-0.043***
	(0.017)	(0.018)	(0.015)	(0.008)
Financial Literacy	-0.138***	0.062**	0.054*	0.021
	(0.030)	(0.029)	(0.030)	(0.016)
Financial Confidence	-0.020***	-0.015**	0.033***	0.020***
	(0.007)	(0.006)	(0.007)	(0.004)
N	7777	7777	7777	7777
Pseudo R^2	5.05%	2.08%	8.23%	8.78%

Notes: Robust standard errors are shown in parentheses. *,**,*** represent statistical significance at the 10%, 5%, and 1% levels respectively. Marginal effects from estimating Equation I using on the choice of using the stimulus check to pay *Bills* (I), pay down *Debt* (II), add to *Saving* (III) or invest in the *Stock* market (IV). Sample includes only salaried workers who received a stimulus check as part of the CARES package. Data are from the 2021 iteration of the National Financial Capability Study.

Table 4: Probit Results Including Knowledge and Confidence Groups

	Bills (I)	Debt (II)	Saving (III)	Stock (IV)
Female	-0.0011	0.0333**	0.0058	-0.0637***
	(0.0141)	(0.0137)	(0.0144)	(0.0075)
Black	0.0362*	0.0241	-0.0199	0.0386***
	(0.0218)	(0.0214)	(0.0222)	(0.0138)
Hispanic	0.0116	0.0114	-0.0151	0.0312**
	(0.0232)	(0.0226)	(0.0240)	(0.0149)
Asian	-0.0636**	-0.0613**	0.0401	0.0276
	(0.0313)	(0.0296)	(0.0315)	(0.0194)
Other Ethnicity	0.0544	-0.0006	0.0092	0.0314
	(0.0368)	(0.0354)	(0.0373)	(0.0245)
Less than HS	0.0118	-0.0053	-0.0753	0.0026
	(0.0669)	(0.0683)	(0.0691)	(0.0457)
Some College	0.0498**	0.0556***	0.0203	0.0330**
	(0.0194)	(0.0191)	(0.0203)	(0.0129)
College	0.0083	0.0289	0.0855***	0.0748***
	(0.0209)	(0.0205)	(0.0216)	(0.0154)
Graduate	0.0098	0.0056	0.1411***	0.0873***
	(0.0262)	(0.0257)	(0.0272)	(0.0212)
Married	0.0048	0.0126	-0.0028	-0.0301***
	(0.0159)	(0.0154)	(0.0160)	(0.0087)
Children	0.0346***	0.0396***	-0.0283***	0.0093***
	(0.0066)	(0.0062)	(0.0067)	(0.0032)
Income 2	-0.0973***	0.0420**	0.0232	0.0160
	(0.0178)	(0.0172)	(0.0181)	(0.0102)
Income 3	-0.1556***	-0.0075	0.0259	0.0242*
	(0.0226)	(0.0221)	(0.0228)	(0.0133)
Income 4	-0.2158***	-0.1175***	0.0723	0.1656***
	(0.0489)	(0.0448)	(0.0544)	(0.0462)
Financial Fragile	0.1405***	0.1081***	-0.2924***	-0.0430***
	(0.0167)	(0.0172)	(0.0149)	(0.0082)
High Knowledge-High Confidence	-0.1226***	-0.0040	0.0817***	0.0554***
	(0.0202)	(0.0198)	(0.0208)	(0.0143)
High Knowledge-Low Confidence	-0.0599***	0.0408**	0.0250	0.0223*
	(0.0193)	(0.0187)	(0.0195)	(0.0122)
Low Knowledge-High Confidence	-0.0138	-0.0413**	0.0318	0.0582***
	(0.0209)	(0.0196)	(0.0210)	(0.0150)
N	7884	7884	7884	7884
Pseudo R2	5.09%	2.21%	8.06%	8.76%

Notes: Robust standard errors are shown in parentheses. *,**,*** represent statistical significance at the 10%, 5%, and 1% levels respectively. Marginal effects from estimating Equation I using on the choice of using the stimulus check to pay *Bills* (I), pay down *Debt* (II), add to *Saving* (III) or invest in the *Stock* market (IV). Sample includes only salaried workers who received a stimulus check as part of the CARES package. Data are from the 2021 iteration of the National Financial Capability Study.

Also, Column III of Table 4 suggests that it is the combined effect of knowledge and confidence that is associated with a greater likelihood of using the EIP to save. Only the marginal effect on Group I is statistically significant, meaning that the likelihood of using the stimulus check to increase savings is not different between the underconfident, the overconfident, and those with low-knowledge and low-confidence. However, Group I, the knowledgeable and confident group, exhibit an 8.17-percentage point higher likelihood of deploying their stimulus payment toward savings than those in Group IV.

Finally, Column IV suggests that confidence is the primary, but not the only, influence for using the stimulus payment to invest in the stock market; Groups I and III, the high confidence groups, show statistically significant marginal effects of similar magnitude. Respondents who are highly confident, irrespective of knowledge, show a 5-percentage point higher likelihood of deploying their EIPs to the stock market. However, when we consider Group II, if confidence is held at low levels, having high knowledge is associated with only a 2.23 percentage point higher likelihood of investing in the stock market.

Overall, the empirical results show that financial knowledge and financial confidence exert distinct and sometimes opposing influences on households' allocation of stimulus payments in times of significant economic uncertainty. Greater knowledge consistently reduces reliance on EIPs for short-term consumption and increases the likelihood of debt repayment and additional savings. Confidence, by contrast, appears to motivate risk-taking: it is positively related to stock market participation but negatively related to debt repayment. Furthermore, the analysis indicates that the overconfident are not less likely to use their EIPs to fund consumption, but they are less likely to reduce debt. In contrast, respondents with high knowledge and high confidence are the most likely to save. These results highlight the importance of distinguishing between financial literacy and financial confidence to enhance the understanding of the drivers behind financial decision-making.

Conclusion

This paper examines if and how financial literacy and confidence relate to households' allocation of economic impact payments received during the COVID-19 pandemic. Using data from the 2021 National Financial Capability Study and probit models for four distinct uses of the EIPs—consumption, debt reduction, additional saving and stock market investment—this study finds further support for the importance of both financial literacy and confidence in financial decision-making. Specifically, higher financial literacy is associated with a lower likelihood of using the EIP to fund consumption, but a greater likelihood of deploying the EIP toward debt reduction and additional savings. In addition, the findings in this paper suggest that financial confidence is also related to the decisions of how to deploy the stimulus check, but this relationship is distinct from financial knowledge. Specifically, higher levels of confidence are consistent with a greater likelihood that the EIP is deployed toward stock market investment, but a lower likelihood that it will be used for debt reduction. Additionally, the results indicate that overconfident respondents are particularly prone to risky behaviors—exhibiting lower likelihoods for debt reduction, but higher likelihoods for stock market investment.

This paper suffers from two important limitations. First, the results are correlational as the cross-sectional design does not allow for causal interpretation. Second, the survey data captures only whether respondents allocated funds to different categories, not the percentage of funds allocated to each. Therefore, we can only document participation in categories and not intensity of allocation. Future studies could build on this study by examining how households deploy liquidity over time, or by examining whether the effects of literacy and confidence persist across subsequent rounds of fiscal stimulus.

Despite these limitations, the paper's findings pose important policy implications. Fiscal stimulus packages aimed at maximizing consumption should carefully consider household characteristics, such as financial literacy and confidence, which shape whether funds are spent, saved, or invested. In addition, financial education interventions should aim to enhance financial knowledge but not neglect the role of confidence for decision-making particularly as it relates to credit management.

Overall, the results highlight the importance of distinguishing between financial literacy and confidence for analyzing household behavior when confronted with an unexpected liquidity infusion. Literacy seems to steer individuals toward saving and debt repayment while confidence can motivate risk-taking, particularly among the overconfident. Recognizing this distinction is critical for understanding financial decision-making and for designing programs that promote well-being in times of economic uncertainty.

Notes

1. See Ouachani, Belhassine, and Kammoun (2020) for a review of the financial literacy literature.
2. We also measure liquidity constraints using two additional measures: The NFCS asks, "In a typical month, how difficult is it for you to cover your expenses and pay all your bills?" We define individuals as *liquidity-constrained* if they respond to this question with "Somewhat Difficult" or "Very Difficult". If they answer, "Not at all Difficult", the binary variable is coded as zero. To account for the possibility that an individual may have set aside money for particularly precarious situations despite finding it difficult to pay their expenses every month, we employ an additional variable. Specifically, the NFCS asks individuals: "*Have you set aside emergency or rainy day funds that would cover your expenses for 3 months, in case of sickness, job loss, economic downturn, or other emergencies?* We create the binary variable *emergency fund* that takes on a value of 1 if the individual responds "Yes" to the aforementioned question and zero otherwise. Results tables show only the financial fragility variable, but empirical results using other liquidity measures are qualitatively similar and available from the author upon request.
3. A major limitation of the NFCS data is that one cannot see the percentage of funds allocated to the different purposes. Yet, as shown in Table 1, more than 40% of respondents indicated spreading their stimulus checks across more than one use. Therefore, as a robustness check, we assess the role of financial literacy and confidence on allocations for respondents that chose only one use. The results, available from the author upon request, are weaker in magnitude, but qualitatively similar to those in Table 4.

References

Allgood S, Walstad WB (2013) Financial literacy and credit card behaviors: A cross-sectional analysis by age. Numeracy 6(2):1-26

Allgood S, Walstad WB (2015) The effects of perceived and actual financial literacy on financial behaviors. Economic Inquiry. DOI: 10.1111/ecin.12255

Armantier O, Goldman L, Koşar G, Lu J, Pomerantz R, van der Klaauw W (2020) How have households used their stimulus payments and how would they spend the next? Liberty Street Economics

Asaad CT (2015) Financial literacy and financial behavior: Assessing knowledge and confidence. Financial Services Review 24(1):101-117

Atlas SA, Lu J, Micu PD, Porto N (2019) Financial knowledge, confidence, credit use, and financial satisfaction. Journal of Financial Counseling and Planning 30(2):175-190

Baker SR, Farrokhnia RA, Meyer S, Pagel M, Yannelis C (2020) Income, liquidity, and the consumption response to the 2020 economic stimulus payments. NBER Working Paper No. 27097

Bhutta N, Blair J, Dettling LJ, Moore KB (2020) COVID-19, the CARES Act and families' financial security https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3631903

Bucher-Koenen T, Lusardi A, Alessie R, van Rooij M (2016) How financially literate are women? An overview and new insights. Journal of Consumer Affairs 51(2):255-283

Coibion O, Gorodnichenko Y, Weber M (2020) How did US consumers use their stimulus payments? NBER Working Paper No. 27693

Cox N, Deadman E, Farrell D, Ganong P, Greig F, Noel P, Vavra J, Wong A (2020) Initial impacts of the pandemic on consumer behavior. Brookings Papers on Economic Activity https://www.jstor.org/stable/10.2307/26996635

Falcettoni E, Nygaard VM (2021) A literature review on the impact of increased unemployment insurance benefits and stimulus checks in the United States. Covid Economics 64(13):168-201

FINRA Investor Education Foundation (2021) National Financial Capability Study

Greenwood R, Laarits T, Wurgler J (2022) Stock market stimulus. NBER Working Paper No. 29827

Hastings JS, Madrian BC, Skimmyhorn WL (2012) Financial literacy, financial education, and economic outcomes. NBER Working Paper No. 18412

Kalenkoski CM, Pabilonia SW (2022) Impacts of COVID-19 on the self-employed. Small Business Economics 58(2):741-768

Kantamneni N (2020) The impact of the COVID-19 pandemic on marginalized populations in the United States: A research agenda. Journal of Vocational Behavior 119:103439

Karpman M, Acs G (2020) Unemployment insurance and economic impact payments associated with reduced hardship following CARES Act. Urban Institute https://www.urban.org/research/publication/unemployment-insurance-and-economic-impact-payments-associated-reduced-hardship-following-cares-act

Karger E, Rajan A (2020) Heterogeneity in the marginal propensity to consume: Evidence from COVID-19 stimulus payments. Chicago Fed Working Paper No. 2020-15

Kim KT, Lee J, Hanna SD (2020) The effects of financial literacy overconfidence on the mortgage delinquency of US households. Journal of Consumer Affairs 54(2):517-540

Klapper LF, Lusardi A, Panos GA (2012) Financial literacy and the financial crisis. World Bank Policy Research Working Paper No. WPS 5980

Kobayashi S, Nakahara K, Oda T, Ueno Y (2020) The impact of COVID-19 on US consumer spending: Quantitative analysis using high-frequency state-level data. Bank of Japan Review https://www.boj.or.jp/en/research/wps_rev/rev_2020/rev20e07.htm

Kyrychenko V, Shum P (2009) Who holds foreign stocks and bonds? Characteristics of active investors in foreign securities. Financial Services Review 18(1):1-21

Levine R, Rubinstein Y (2017) Smart and illicit: Who becomes an entrepreneur and do they earn more? Quarterly Journal of Economics 132(2):963-1018

Lu Y, Kalenkoski CM (2023) Objective financial knowledge, subjective financial knowledge, and credit card mismanagement: Does overconfidence hurt? Journal of Financial Counseling and Planning 34(3):430-444

Lusardi A (2008) Financial literacy: An essential tool for informed consumer choice? NBER Working Paper No. 14084

Lusardi A (2012) Numeracy, financial literacy, and financial decision-making. NBER Working Paper No. 17821

Lusardi A, de Bassa Scheresberg C (2013) Financial literacy and high-cost borrowing in the United States. NBER Working Paper No. 18969

Lusardi A, Mitchell OS (2007) Financial literacy and retirement preparedness: Evidence and implications for financial education. Business Economics 42(1):35-44

Lusardi A, Mitchell OS (2008) Planning and financial literacy: How do women fare? NBER Working Paper No. 13750

Lusardi A, Mitchell OS (2011a) Financial literacy and planning: Implications for retirement wellbeing. NBER Working Paper No. 17078
Lusardi A, Mitchell OS (2011b) Financial literacy and retirement planning in the United States. Journal of Pension Economics and Finance 10(4):509-525
Lusardi A, Mitchell OS (2014) The economic importance of financial literacy: Theory and evidence. NBER Working Paper No. 18952
Lusardi A, Schneider DJ, Tufano P (2011) Financially fragile households: Evidence and implications. NBER Working Paper No. 17072
Mishra DSK (2023) Demonetisation: A blessing for digitalisation. SSRN
Misra K, Singh V, Zhang Q (2022) Frontiers: Impact of stay-at-home-orders and cost-of-living on stimulus response: Evidence from CARES Act. Marketing Science 41(2): 211-229
Ouachani S, Belhassine O, Kammoun A (2020) Measuring financial literacy: A literature review. Managerial Finance 47(2):266-281
Parker AM, Bruine de Bruin W, Yoong J, Willis R (2012) Inappropriate confidence and retirement planning: Four studies with a national sample. Journal of Behavioral Decision Making 25(4):382-389
Robb CA, Babiarz P, Woodyard A (2012) The demand for financial professionals' advice: The role of financial knowledge, satisfaction and confidence. Financial Services Review 21(4):291-305
Sajid M, Mushtaq R, Murtaza G, Yahiaoui D, Pereira V (2024) Financial literacy, confidence and well-being: The mediating role of financial behavior. Journal of Business Research 182:114791
Toczynski J (2022) Spend or invest? Analyzing MPC heterogeneity across three stimulus programs. Swiss Finance Institute Research Paper No. 32-57 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4270518
Van Rooij M, Lusardi A, Alessie R (2007) Financial literacy and stock market participation. NBER Working Paper No. 13565
Zimmerman P, Divakaruni A (2021) Uncovering retail trading in bitcoin: The impact of COVID-19 stimulus checks. Cleveland FRB of Cleveland Working Paper No. 21-13 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3888393

Appendix

Questions for Constructing Financial Literacy & Confidence Measures

Financial Literacy Question (% Correct)	Possible Answers
Suppose you had $100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow?	More than $102; Exactly $102; Less than $102; Don't know; Prefer not to say
Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, how much would you be able to buy with the money in the account?	More than today; Exactly the same; Less than today; Don't know; Prefer Not To Say
Buying a single company's stock usually provides a safer return than a stock mutual fund.	True; False; Don't know; Prefer not to say
A 15-year mortgage typically requires higher monthly payments than a 30-year mortgage, but the total interest paid over the life of the loan will be less.	True; False; Don't know; Prefer not to say
If interest rates rise, what will typically happen to bond prices?	They will rise; They will fall; They will stay the same; There is no relationship between bond prices and the interest rate; Don't know; Prefer not to say
Suppose you owe $1,000 on a loan and the interest rate is 20% per year compounded annually. If you didn't pay anything off, at this interest rate, how many years would it take for the amount you owe to double?	Less than 2 years; At least 2 years but less than 5 years; At least 5 years but less than 10 years; At least 10 years; Don't know; Prefer not to say

Which of the following indicates the highest probability of getting a particular disease?	There is a one-in-twenty chance of getting the disease; 2% of the population will get the disease; 25 out of every 1,000 people will get the disease; Don't know; Prefer not to say
Financial Confidence How strongly do you agree with the following statements? (1-7 Scale)	
I am good at dealing with day-to-day financial matters, such as checking accounts, credit and debit cards, and tracking expenses.	1-Strongly Disagree; 4-Neither Agree nor Disagree; 7-Strongly Agree; 98-Don't know; 99-Prefer not to say
I am pretty good at math.	1-Strongly Disagree; 4-Neither Agree nor Disagree; 7-Strongly Agree; 98-Don't know; 99-Prefer not to say
On a scale from 1 to 7, how would you assess your overall financial knowledge?	1-Very Low; 7-Very High; 98-Don't know; 99-Prefer not to say

Do Economic Fundamentals Improve Exchange Rate Forecasting? Evidence from Machine Learning Models

Gokcen Ogruk-Maz and Sinan Yildirim, Texas Wesleyan University

Abstract

This paper evaluates whether macroeconomic fundamentals improve short-horizon exchange rate forecasts when combined with modern machine learning models. Using monthly data from 2010 to 2023 for five major currency pairs, we estimate Facebook Prophet, Artificial Neural Networks, and XGBoost with and without fundamentals such as interest rates, inflation, money supply, and output. Forecast accuracy is assessed using the Diebold-Mariano test and Theil's U. We find that fundamentals do not consistently enhance one-month-ahead forecasts. Prophet fails to outperform ARIMA, ANN improves accuracy only occasionally, and XGBoost's strong results likely reflect overfitting. These outcomes align with Meese and Rogoff's conclusion that short-run exchange rates remain difficult to predict.

JEL Codes: F31, C45
Keywords: Exchange Rates, Machine Learning, Forecasting, Economic Fundamentals

Introduction

Forecasting exchange rates continues to be one of the most difficult tasks in international economics. Exchange rates influence international trade, cross border investment, inflation, interest rates, and financial stability. They affect the decisions of households, firms, and policymakers. A change in the value of a currency can alter the cost of imports, the competitiveness of exports, and the overall direction of monetary policy. Because of these effects, we continue to study how exchange rates move over time and whether their movements can be predicted with reasonable accuracy.

Although exchange rates are important, academic research shows that forecasting them is very challenging. The study by Meese and Rogoff (1983) showed that well-known structural exchange rate models did not perform better than a simple random walk in out-of-sample forecasting. Their study compared models based on Purchasing Power Parity, Uncovered Interest Rate Parity, and monetary fundamentals. These models are grounded in well-established economic theory. The surprising result that they do not outperform a random walk became known as the exchange rate disconnect puzzle. The puzzle suggests that exchange rates behave almost like random processes in the short run and that traditional fundamentals do not provide a strong basis for short-horizon forecasting.

Later studies tried to revisit this conclusion. Mark (1995) argued that fundamentals may improve forecasts at longer horizons. Molodtsova and Papell (2009) showed that Taylor Rule fundamentals can provide improvements relative to the random walk in some cases. Chen (2011) suggested that fundamentals may help predict the direction of exchange rate movements and that survey-based expectations may include information that standard models ignore. These studies expanded the discussion on when fundamentals matter. However, the main conclusion in the literature remains that forecasting at the one-month horizon is difficult and that fundamentals often provide limited help.

Developments in computational methods encouraged researchers to explore more flexible forecasting tools. Machine learning methods became popular in economics and finance because they can capture nonlinear relationships and complex interactions that traditional econometric models may miss. Artificial Neural Networks are frequently used in studies of stock markets, inflation, and exchange rates, including applications by Garlapati et al. (2021) and other studies in financial forecasting. These models can learn patterns in the data without a predetermined structure.

Facebook Prophet has also received attention in applied research. It models changes in trends and seasonality in a simple and practical way. Prophet is used in business analytics and in many applied forecasting contexts. Studies include supermarket sales forecasting by Jha and Pande (2021), oil price forecasting by Güleryüz and Özden (2020), stock price forecasting by Garlapati et al. (2021), sales forecasting by Zunic et al. (2020), and public health forecasting by Gaur et al. (2020). These studies show that Prophet can perform well in applied work. However, Prophet has not been examined in a systematic way in foreign exchange markets where volatility is high and seasonality is limited.

Other machine learning methods, such as XGBoost and Random Forests have also been used in financial forecasting. Pfahler (2022) reported strong results for several exchange rates. Tsuji (2022) found that machine learning models performed well in short-horizon forecasting. Pryimak et al. (2023) reached similar conclusions in their study of the Ukrainian hryvnia and emphasized the importance of careful validation because machine learning models may overfit when the data contain a high amount of noise.

Although the machine learning literature continues to grow, important gaps remain. Many studies focus on a single model or on a single currency pair. This makes it difficult to compare performance across currencies and across forecasting methods. In addition, even though fundamentals such as interest rates, inflation, money supply, and output gaps are central in theoretical exchange rate models, only a small number of studies ask whether these fundamentals improve machine learning forecasts in practice. Prophet is widely used in applied work, yet we do not know how it performs in the foreign exchange market.

In this paper, we address these gaps by applying three forecasting models, Facebook Prophet, Artificial Neural Networks, and XGBoost, to five major currency pairs using monthly data from 2010 to 2023. For each model, we estimate a version that uses only past exchange rate information and a version that also includes fundamentals from Uncovered Interest Rate Parity, Purchasing Power Parity, the Monetary Model, and the Taylor Rule. These variables allow us to test whether fundamentals can improve predictive accuracy when used within modern forecasting frameworks. We evaluate forecast accuracy using the Diebold-Mariano test from Diebold and Mariano (1995) and the Theil U statistic from Theil (1966).

By comparing several forecasting models with and without fundamentals, we provide new evidence on whether macroeconomic variables help predict exchange rate movements at the one-month horizon. Our results show that fundamentals do not consistently improve forecasts and that the strongest performance by XGBoost is more consistent with overfitting than with genuine predictive content. These findings agree with the perspective of Meese and Rogoff (1983). Even when we use modern machine learning tools, fundamentals do not provide systematic gains in short-horizon forecasting. Our study, therefore, contributes to the forecasting literature and to the continuing discussion on the relationship between fundamentals and exchange rate dynamics.

Methodology

This section presents the methodological framework used in this study. We begin by describing the four economic fundamental models that provide the theoretical foundation for exchange rate determination. These models are Uncovered Interest Rate Parity, Purchasing Power parity, the Monetary Model, and the Taylor Rule, and they form the basis for constructing explanatory variables. Following this, we introduce the three machine learning models applied in our forecasting exercises: Facebook Prophet, Artificial Neural Networks, and XGBoost. Each model is evaluated both with and without the inclusion of fundamental variables. Finally, we explain the statistical procedures used to evaluate forecast accuracy, including the Diebold-Mariano test and Theil's U statistic, which allow for comparison of model performance relative to benchmarks and each other.

Economic Fundamentals Models

To evaluate the role of macroeconomic factors in exchange rate forecasting, we incorporate four established economic models. These models are selected based on their strong theoretical foundations in international finance and their relevance in previous forecasting research. Each model captures a different mechanism by which exchange rates respond to changes in interest rates, inflation, money supply, or output. By using these models, we aim to examine whether combining these traditional indicators with machine learning algorithms enhances forecasting accuracy.

Uncovered Interest Rate Parity (UIP) assumes that the change in the exchange rate equals the interest rate differential between two countries:

$$\Delta s_t = r_{d,t} - r_{f,t} \tag{1}$$

where, Δs_t is the change in the log of the exchange rate, $r_{d,t}$ is the domestic interest rate, and $r_{f,t}$ is the foreign interest rate. According to UIP, investors should earn the same return on investments in different currencies, once exchange rate changes are accounted for.

Purchasing Power Parity (PPP) is based on price level differences across countries and assumes that exchange rates adjust to reflect inflation gaps:

$$\Delta s_t = \pi_{d,t} - \pi_{f,t} \tag{2}$$

where $\pi_{d,t}$ and $\pi_{f,t}$ are the inflation rates in the domestic and foreign country, respectively. PPP suggests that over time, goods should cost the same in different countries once currency differences are adjusted.

The Monetary Model (MM) assumes that exchange rates respond to differences in monetary and real economic conditions between countries. Specifically, the model suggests that currencies with higher money supply growth or lower output growth relative to their foreign counterparts will tend to depreciate. The version used in this study follows the specification adopted by Amat et al. (2018), which models exchange rate changes as a linear function of changes in money supply and output levels across countries:

$$\Delta s_t = b_1 \Delta m_{d,t} - b_2 \Delta m_{f,t} - b_3 \Delta y_{d,t} - b_4 \Delta y_{f,t} \tag{3}$$

where m represents log money supply and y represents log output (income). The model assumes that currencies with higher money growth or lower real output growth tend to depreciate.

The Taylor Rule Model combines inflation, output, and interest rate differences. It is based on how central banks set interest rates in response to economic conditions. The model suggests that currencies from countries with higher inflation, faster growth, or tighter monetary policy should appreciate.

$$\Delta s_t = (\pi_{d,t} - \pi_{f,t}) + (y_{d,t} - y_{f,t}) + r_{d,t} - r_{f,t} \tag{4}$$

Machine Learning Models

We use three machine learning models: Facebook Prophet, Artificial Neural Networks, and XGBoost. Each model is applied to forecast exchange rates with and without economic fundamentals.

Facebook Prophet is a modular additive time series forecasting model developed by the company Meta. It is designed to handle time series data with strong seasonal patterns and potential trend changes. The model decomposes the series into the following components: trend, seasonality, holiday effects, and residuals.

$$y_t = g_t + s_t + h_t + x_t + \epsilon_t \tag{5}$$

where g_t represents the non-periodic trend component, s_t captures seasonality using a Fourier series, h_t reflects holiday effects, x_t denotes the effect of external regressors, and ϵ_t is the error term. The model supports linear and logistic growth functions with automatic changepoint detection. In this study, we extend the Prophet model by including fundamental macroeconomic indicators such as interest rate, inflation, output, and money supply as external regressors in the x_t component. This allows the model to incorporate the time series pattern together with the economic variables as forecasting inputs. We use these variables only to evaluate whether they improve predictive accuracy.

Artificial Neural Networks (ANN) are a class of machine learning models inspired by the structure of the human brain. They consist of layers of interconnected nodes, known as neurons, which process and transmit information. In a standard feedforward ANN, information flows from the input layer through one or more hidden layers to the output layer. Each neuron applies a linear transformation followed by a nonlinear activation function, such as the Rectified Linear Unit (ReLU) or the sigmoid function, allowing the network to learn complex and nonlinear patterns in the data.

In this study, we implement a feedforward ANN with one hidden layer and a small number of neurons to reduce overfitting. The model is trained using monthly exchange rate data and macroeconomic inputs. We evaluate performance for each currency pair both with and without the inclusion of economic fundamentals to test their contribution to predictive accuracy. The functional form of a single-layer ANN is:

$$y = f(Wx + b + Z\beta) \tag{6}$$

where W represents the matrix of weights for time series inputs x, b is the bias term, and $Z\beta$ captures the effect of macroeconomic inputs. The vector Z contains the fundamental variables and β represents their associated weights. The activation function introduces nonlinearity and is typically chosen as ReLU or sigmoid. The model is trained using backpropagation, adjusting weights to minimize forecast error.

XGBoost (Extreme Gradient Boosting) is an ensemble learning method based on sequentially constructed decision trees. In each round, a new tree is added to minimize the residual errors of the previous model, improving the forecast in an iterative way. In this study, we use XGBoost to forecast monthly exchange rates with both lagged exchange rates and macroeconomic indicators as input features. The model can work with a large number of inputs and capture nonlinear patterns in the data in a flexible forecasting setting. The prediction for an observation i is modeled as:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i, z_i), \qquad f_k \in \mathcal{F} \tag{7}$$

where x_i includes lagged exchange rate terms and z_i includes macroeconomic inputs such as interest rate, inflation, money supply, and output. Each function f_k represents a regression tree. The model minimizes a regularized objective function:

$$L(\phi) = \sum_i l(\widehat{y_i}, y_i) + \sum_k \Omega(f_k) \tag{8}$$

with the regularization term defined as:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda\|w\|^2 \tag{9}$$

where T is the number of leaves in a tree, w represents the vector of leaf weights, γ and λ are regularization parameters to control model complexity and prevent overfitting.

Evaluation Metrics

To evaluate the performance of forecasting models, we use two widely accepted statistical tools: the Diebold-Mariano (DM) test and Theil's U statistic. These measures assess both the relative accuracy between competing models and the absolute performance of a model compared to a benchmark model, with the random walk model, which serves as the benchmark in this study

Diebold-Mariano Test is used to determine whether the forecast errors from two models are significantly different from one another. It calculates the average difference in forecast errors over time, adjusting for serial correlation in the data. A positive and statistically significant result suggests that one model consistently outperforms the other. This test is particularly suitable for our study, as it helps evaluate whether machine learning models with or without macroeconomic fundamentals lead to a meaningful improvement in forecast accuracy.

Theil's U provides a ratio that compares the forecast accuracy of a given model to that of a naive benchmark. It is calculated using the root mean squared error (RMSE) from the forecast model and the RMSE from the benchmark model. A value less than one indicates better performance than the benchmark. This metric was chosen because it gives a simple and interpretable measure of whether a model adds forecasting value beyond simply repeating the last known observation.

These two metrics were selected for their complementary strengths. The DM test is applied to compare the forecast errors of each machine learning model and those of the random walk benchmark. This allows us to test whether the model significantly outperforms the random walk forecast, and whether the addition of theoretical economic variables improves forecast performance. The Theil's U statistic is computed for each model and currency pair, allowing us to benchmark model performance against a benchmark ARIMA model.

Data and Empirical Results

This study uses monthly data from January 2010 to December 2023 for five major currency pairs: JPY/USD, AUD/USD, CAD/USD, GBP/USD, and NZD/USD. We begin the sample in 2010 in order to work within a consistent monetary and financial environment. The period before 2010 includes several major structural breaks such as the Asian Financial Crisis, the dot-com recession, and the Global Financial Crisis. These events introduced lasting changes in exchange rate behavior and in the links between exchange rates and macroeconomic fundamentals. Machine learning models tend to treat such regime changes as part of the underlying pattern, which reduces the reliability of the forecast evaluation. The post-2010 period reflects a more coherent monetary policy framework in which inflation targeting, policy communication, and interest rate adjustment follow a more uniform structure across countries. Although the COVID shock created temporary volatility, it did not create a permanent regime shift in exchange rate determination. For these reasons, the period after 2010 provides a consistent setting for examining whether fundamentals improve short-horizon exchange rate forecasts.

Exchange rate data are obtained from the Federal Reserve Economic Data (FRED) database, where each exchange rate is expressed as the price of the foreign currency in terms of U.S. dollars (USD). All exchange rates are converted to natural logarithms before analysis to ensure stationarity and interpretability of percentage changes. Macroeconomic variables used in the fundamental models include interest rates, consumer price indices (CPI), real gross domestic product (GDP), and money supply (M2). Interest rate and inflation data are sourced from the International Monetary Fund (IMF) and the World Bank. Output and money supply data are retrieved from the OECD database. All macroeconomic series are seasonally adjusted and transformed into natural logs before differencing, following common practice in empirical exchange rate literature. The data are divided into two subperiods: the training period from January 2010 to December 2020 and the test period from January 2021 to December 2023. Forecasts are made one month ahead using rolling window estimation, allowing updated parameter estimates at each step.

Table 1 reports the one-month-ahead forecasting results using the Facebook Prophet model, with and without the inclusion of fundamental variables, across five currency pairs. In all comparisons, the benchmark model is ARIMA. The ARIMA specification is selected with the automatic procedure by Hyndman and Khandakar (2008) using the AIC criterion. This method tests for unit roots, estimates the model by maximum likelihood, and selects the order that gives the best statistical fit. The ARIMA model is re-estimated at every forecasting point because the sample size grows over time. This approach reflects real-time forecasting conditions and follows the recommendation in Rossi (2013). Using ARIMA as the benchmark is standard in the exchange rate literature. Meese and Rogoff (1983) show that ARIMA is a difficult benchmark to beat in out-of-sample forecasts. Later studies, such as Cheung Chinn and Pascual (2005) confirm that ARIMA remains a useful comparison model when evaluating forecasts from economic fundamentals or machine learning methods.

The table presents the DM test statistics, p-values, and Theil ratios for each specification. A negative DM statistic with a p-value below 0.05 suggests that the given model performs significantly better than the benchmark random walk model. Theil ratios below 1 indicate improved forecast accuracy relative to the benchmark. Results show that the baseline Facebook Prophet model does not outperform the ARIMA benchmark, as all Theil ratios are above 1 and the DM test results are not statistically significant for any currency pair. When macroeconomic fundamentals are added, forecasting performance does not consistently improve. In most cases, Theil ratios remain close to or above 1, indicating no systematic gain over the baseline model. For example, the UIP, PPP, and Taylor Rule specifications generally produce similar or slightly higher forecasting errors compared to Prophet alone, and none of the DM statistics indicate significant improvement across currencies. Some mixed results appear, such as slightly lower Theil ratios for JPY/USD under the Monetary Model or for CAD/USD under the Taylor Rule, but these improvements are small and statistically insignificant. The addition of fundamentals tends to yield currency- and model-specific changes rather than broad or consistent enhancements in predictive accuracy.

Table 1. One-Month-Ahead Forecasting of Exchange Rate Differentials with Fundamentals Using Facebook Prophet

Model	**Statistic**	**JPY/USD**	**AUD/USD**	**CAD/USD**	**GBP/USD**	**NZD/USD**
Facebook Prophet	DM	0.3404	0.4684	0.4094	0.8839	0.2480
	p-value	0.7347	0.6411	0.6835	0.3801	0.8049
	Theil	1.0178	1.0255	1.0223	1.0245	1.0077
Prophet + UIP	DM	0.1143	–0.0348	–0.1948	*NA*	0.1244
	p-value	0.9093	0.9723	0.8461	*NA*	0.9013
	Theil	1.0070	0.9982	0.9905	*NA*	1.0034
Prophet + PPP	DM	0.3406	0.6403	-0.0289	−0.4381	–0.2005
	p-value	0.7345	0.5244	0.9770	0.6629	0.8417
	Theil	1.0166	1.0372	0.9987	0.9834	0.9915
Prophet + MM	DM	-0.2495	1.4854	1.1915	1.6622	*NA*
	p-value	0.8038	0.1433	0.2386	0.1022	*NA*
	Theil	0.9866	1.1823	1.0985	1.1573	*NA*
Prophet + Taylor Rule	DM	1.2495	0.3492	−0.936	1.2495	0.1542
	p-value	0.2168	0.7283	0.3529	0.2168	0.8780
	Theil	0.9905	1.0205	0.9438	1.0364	1.0052
Prophet + All Fundamentals	DM	0.2855	1.0349	0.0517	*NA*	*NA*
	p-value	0.7763	0.3053	0.9589	*NA*	*NA*
	Theil	1.0183	1.0555	1.0028	*NA*	*NA*

Table 2 presents the one-month-ahead forecasting results from the ANN model, comparing forecasts with and without economic fundamentals across five currency pairs. As in the previous table, ARIMA is used as the benchmark model. Results indicate that the baseline ANN model without fundamentals performs reasonably well, with Theil ratios below 1 for four out of five currency pairs, although none of the DM test results are statistically significant. Adding fundamentals improves performance in one case only, particularly when PPP is included for CAD/USD, where the Theil ratio drops well below 1 and the DM statistic becomes significant. The Taylor Rule also produces modest gains for NZD/USD, although the DM values are not statistically significant. However, these improvements are not consistent across all currencies or specifications. The CAD/USD pair shows the clearest improvement with fundamentals, while for others like GBP/USD, the gains are weaker or not statistically meaningful. These findings suggest that the ANN model performs well on its own, and in one case, adding fundamentals such as PPP can help improve forecast accuracy. However, this improvement is not seen in every currency pair, so we cannot say that adding fundamentals always leads to better results.

Table 3 presents the one-month-ahead forecasting results using the XGBoost model, comparing its performance with and without the inclusion of economic fundamentals across five currency pairs. The ARIMA model is again used as the benchmark. The baseline XGBoost model outperforms all other specifications by a wide margin, with highly significant DM test results (p-values = 0.0000) and very low Theil ratios across all five currency pairs. This result aligns with the rolling, expanding-window design in the code, where we re-estimate XGBoost each period using lagged monetary and industrial production growth and then compare one-step-ahead log-return forecasts with ARIMA using RMSE, Theil's U, and the Diebold–Mariano test. We already incorporate several regularization features, including a reduced learning rate that slows how much each new tree changes the model, random sampling of observations and predictors when we grow each tree, limits on how deep the trees can grow, and an L2 penalty that shrinks large leaf weights, which together constrain model complexity in line with established recommendations for boosted tree models (Elith, Leathwick, & Hastie, 2008; Hastie, Tibshirani, & Friedman, 2009).

Table 2. One-Month-Ahead Forecasting of Exchange Rate Differentials with Fundamentals Using Artificial Neural Network

Model	Statistic	JPY/USD	AUD/USD	CAD/USD	GBP/USD	NZD/USD
ANN	DM	-0.2939	-1.0649	0.0622	-0.3100	-0.9394
	p-value	0.7697	0.2910	0.9505	0.7575	0.3511
	Theil	0.9927	0.9607	1.0020	0.9854	0.9493
ANN + UIP	DM	1.0412	-0.5435	0.3316	*NA*	-0.4170
	p-value	0.3018	0.5887	0.7412	*NA*	0.6781
	Theil	1.0608	0.9817	1.0166	*NA*	0.9827
ANN + PPP	DM	0.3406	-1.0310	-2.712	-1.2504	0.4204
	p-value	0.7345	0.3067	**0.0086**	0.2158	0.6756
	Theil	1.0166	0.9592	0.7761	0.9319	1.0272
ANN + MM	DM	0.5407	0.5242	2.8914	-0.6985	*NA*
	p-value	0.5909	0.6022	0.0055	0.4878	*NA*
	Theil	1.0276	1.0327	1.1814	0.9471	*NA*
ANN + Taylor Rule	DM	0.5289	1.6595	1.7901	-0.3923	-1.1853
	p-value	0.5990	0.1029	0.0790	0.6963	0.2411
	Theil	1.0225	1.1964	1.0939	0.9749	0.9307
ANN + All Fund.	DM	0.5349	-0.4290	0.7371	*NA*	*NA*
	p-value	0.5949	0.6696	0.4642	*NA*	*NA*
	Theil	1.0282	0.9753	1.0589	*NA*	*NA*

At the same time, the literature on gradient-boosted trees cautions that exceptionally strong forecasting performance, particularly when models are re-estimated on overlapping windows, may reflect overfitting rather than genuine predictive power (Hastie et al., 2009). Chen and Guestrin (2016) also note that boosting algorithms can easily capture very specific patterns in the training data, which may not extend to new periods, and therefore they stress the need for rigorous out-of-sample testing. This concern is especially relevant in the exchange-rate forecasting literature, where the Meese–Rogoff (1983) results repeatedly show that models with excellent in-sample fit often fail to outperform simple benchmarks once moved to true out-of-sample environments. In our case, the extremely low Theil ratios from the baseline XGBoost model are promising, but such strong results may be difficult to sustain beyond the estimation period. For this reason, we interpret the XGBoost findings with caution and emphasize the importance of future research using rolling or expanding out-of-sample windows to confirm that the observed gains represent real predictive structure rather than overfitting.

Discussion and Conclusion

In this study, we evaluate whether macroeconomic fundamentals improve the short-horizon forecasting performance of three modern forecasting frameworks: Facebook Prophet, Artificial Neural Networks, and XGBoost. We examine five major currency pairs and use ARIMA as the benchmark model, which follows common practice in the exchange rate literature.

Through this approach, we are able to assess the value of fundamentals relative to model architecture and to observe how different forecasting tools behave in a noisy environment.

The results for Facebook Prophet show that the model does not outperform ARIMA. The Theil ratios remain above one for the majority of the currency pairs and the Diebold Mariano statistics do not indicate meaningful gains. When we add fundamentals such as UIP, PPP, the Monetary Model, or the Taylor Rule, accuracy does not improve in a systematic way. These results confirm the difficulty of beating simple benchmarks in short-horizon exchange rate forecasting, a finding that is consistent with the broader literature.

The Artificial Neural Network model performs somewhat better than Prophet. We observe cases where ANN forecasts approach or improve upon the ARIMA benchmark, and PPP produces a clear improvement for CAD when evaluated with both Theil's U and the Diebold Mariano statistic. However, these gains do not generalize to the full set of currencies.

Table 3. One-Month-Ahead Forecasting of Exchange Rate Differentials with Fundamentals Using XGBoost

Model	Statistic	JPY/USD	AUD/USD	CAD/USD	GBP/USD	NZD/USD
XGBoost	DM	-4.4739	-5.5845	-5.6722	-5.1773	-4.7276
	p-value	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
	Theil	0.1041	0.0638	0.0185	0.0248	0.0196
XGBoost + UIP	DM	0.6127	0.4567	2.4174	*NA*	0.2395
	p-value	0.5422	0.3586	0.0186	*NA*	0.8114
	Theil	1.0517	1.0294	1.2134	*NA*	1.0134
XGBoost + PPP	DM	1.4662	0.6971	0.8432	-0.4626	1.5307
	p-value	0.1477	0.4884	0.4024	0.64526	0.1311
	Theil	1.0872	1.0574	1.0533	0.95043	1.0984
XGBoost + MM	DM	1.4662	1.0835	0.8707	0.6401	*NA*
	p-value	0.1477	0.2834	0.3877	0.5247	*NA*
	Theil	1.0872	1.1326	1.1065	1.0421	*NA*
XGBoost + Taylor	DM	0.6171	0.3492	2.3054	0.4923	-0.9712
	p-value	0.5397	0.7283	0.0250	0.5963	0.3359
	Theil	1.0385	1.0206	1.2001	1.0220	0.8959
XGBoost + All Fund.	DM	0.7013	-0.4081	-0.7972	*NA*	*NA*
	p-value	0.4861	0.6848	0.4289	*NA*	*NA*
	Theil	1.0587	0.9645	0.9365	*NA*	*NA*

The XGBoost results require careful interpretation. We find very low forecast errors in the baseline XGBoost model, and the Diebold-Mariano tests show strong differences relative to ARIMA. However, we cannot interpret these results as evidence of superior forecasting performance. XGBoost is a highly flexible learning algorithm, and this flexibility makes overfitting a serious concern. When a model becomes too sensitive to the training sample, it may capture noise rather than meaningful structure. In our setting, the extremely low Theil ratios are more consistent with overfitting than with genuine predictive skill. The fact that forecast accuracy declines sharply once we add fundamentals reinforces this interpretation, because a truly robust model should not deteriorate so quickly when the information set changes. We see these results as a signal that the baseline XGBoost model may be fitting short-run idiosyncrasies rather than stable economic patterns. The exchange rate literature has long warned that strong in-sample performance rarely carries over to real-time forecasting, and our findings fit closely with that experience.

Taken together, our results show that macroeconomic fundamentals do not improve one-month-ahead forecasts across the models we study. The structure of the forecasting model matters much more for accuracy than the inclusion of fundamentals. Additionally, fundamentals do not contribute useful information at this horizon. We find that XGBoost can produce very optimistic in-sample results, but these results may not represent true predictive power. This point reinforces the need to treat highly flexible algorithms with caution in macroeconomic settings where the signal is weak relative to the noise.

A limitation of this study is the relatively short sample period. The period after 2010 provides a consistent policy and financial environment, which is useful for evaluating short-horizon forecasts. However, it also restricts the amount of variation in macroeconomic conditions and limits the range of exchange rate behavior that the models observe during training. A longer sample could include additional episodes of monetary tightening, loosening, and changes in global risk sentiment, which might influence how fundamentals relate to exchange rates at the one-month horizon. Although extending the sample would introduce several large structural breaks, it could still be informative for understanding whether the weak performance of fundamentals is specific to the recent period or reflects a broader pattern. Future research may revisit the analysis with longer or higher frequency datasets or may explore methods that account explicitly for regime changes to study how the predictive content of fundamentals varies across economic conditions.

Future research can expand on our analysis in several ways. One direction is to study how the performance of fundamentals changes across different volatility regimes or monetary policy environments, using methods that identify regime shifts directly. Another direction is to examine higher frequency data, which may capture short-run adjustments in expectations more effectively than monthly series. Extending the set of predictors to include measures of global uncertainty, risk sentiment, and capital flows may also provide new insights into what drives exchange rate movements at short horizons. Additional work that compares machine learning models with structural models in a unified forecasting framework would help clarify when model flexibility adds value and when it becomes a liability. These extensions can contribute to a deeper understanding of how fundamentals interact with forecasting tools in environments where the signal is weak relative to the noise.

References

Amat O, Blasco N, Ferrer R (2018) Forecasting exchange rates with fundamentals and ridge regression. Applied Economics Letters 25(2):91–95

Chen SS (2011) Predicting swings in exchange rates with macro fundamentals. SSRN Electronic Journal

Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 785–794

Chen YC (2011) Exchange rate predictability with macroeconomic fundamentals: Evidence from survey data. Journal of International Economics 84(1):58–77

Cheung YW, Chinn MD, Pascual AG (2005) Empirical exchange rate models of the nineties. Journal of International Money and Finance 24(7):1150–1175

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. Journal of Animal Ecology 77(4):802–813

Garlapati P, Sahoo A, Kumar KR (2021) Stock price prediction using Facebook Prophet and ARIMA models. International Journal of Creative Research Thoughts 9(3):111–118

Güleryüz D, Özden E (2020) The prediction of Brent crude oil trend using LSTM and Facebook Prophet. European Journal of Science and Technology 20:1–9

Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: Data mining, inference, and prediction. Springer

Hyndman RJ, Khandakar Y (2008) Automatic time series forecasting: The forecast package for R. Journal of Statistical Software 27(3):1–22

Jha BK, Pande S (2021) Time series forecasting model for supermarket sales using FB-Prophet. Proceedings of the Fifth International Conference on Computing Methodologies and Communication 546–554

Mark NC (1995) Exchange rates and fundamentals: Evidence on long-horizon predictability. The American Economic Review 85(1):201–218

Meese RA, Rogoff K (1983) Empirical exchange rate models of the seventies: Do they fit out of sample? Journal of International Economics 14(1–2):3–24

Molodtsova T, Papell DH (2009) Out-of-sample exchange rate predictability with Taylor rule fundamentals. Journal of International Economics 77(2):167–180

Pfahler M (2022) Forecasting exchange rates with machine learning: Evidence from XGBoost and artificial neural networks. Journal of Forecasting 41(2):247–263

Pryimak V, Bartkiv B, Holubnyk O (2023) Forecasting the exchange rate of the Ukrainian hryvnia using machine learning methods. Computer Systems and Information Technologies 2023(1):75–82

Rossi B (2013) Exchange rate predictability. Journal of Economic Literature 51(4):1063–1119

Tsuji C (2022) Machine learning-based exchange rate forecasting: Evidence from XGBoost and random forests. Review of Development Economics 26(4):2263–2280

How Elections for Local Property Appraisers Affect Appraised Property Values

Cullen T. Wallace, Georgia College & State University

Abstract

Elected property appraisers may face an incentive to lower or underestimate appraised property values during election season to appeal to voters through lower tax liability. Leveraging data from Florida, I estimate the change in appraised property values during election years compared to non-election years. I find that, on average, appraised values are lower for years preceding an election year, affecting the last tax bill a property owner may pay prior to voting. I also find evidence that property values are undervalued – compared to sales data – prior to an election year, and this effect is greater for counties with challenged incumbents.

JEL Codes: H71, D72
Keywords: Property tax, Local elections

Introduction

Property taxes are the largest source of local-government revenue, aside from state aid. Typically, property taxes are levied annually and involve multiple members of local government: a taxing authority, a property appraiser (i.e. assessor), and a tax collector.[1] Generally, the property appraiser and tax collector follow standard procedure to assess the value of property and collect the taxes due. Entities with the authority to levy taxes (e.g. county commissioners) have the most consequential role in the process – setting millage rates and thereby deciding the actual dollar amount of taxes owed by property owners.

Often, all three positions involved in the property tax are filled using elections, and given a basic understanding of public choice, it is natural that the behavior of these individuals might manifest in such a way as to maximize their chances of retaining their office. Regarding the taxing authority, it is sensible that those bearing the responsibility of imposing taxes would be accountable to voters, and it is fairly straightforward to understand how individuals might compete for these political positions – in a sense, promoting policies with high or low taxes coupled with high or low levels of public goods.

However, given that the assessor and collector generally complete routine administrative tasks, the ways in which these individuals may react to political incentives are less clear. How might the desire to elicit goodwill with potential voters impact the behavior of these office holders – and on what dimensions might these office holders act? This paper addresses a subset of these issues, focusing on the first step of the property tax process, the role of property appraisers and the assessment of property values.

Do appraised property values in jurisdictions with elected property appraisers vary with regard to the election cycle? If so, one might be able to statistically identify an "election-year" effect, in which appraised values are lower during election years than in other years. Appraisers may attempt to suppress property values – and therefore tax liability – in an effort to increase support among voters.

In this paper, I empirically examine twelve years of county-level data on property values in Florida, where appraisers are elected in each county to four-year terms. Using a two-way fixed-effects model, I identify the differential effect of appraised property values in years leading up to an election year. I find evidence that appraised property values are systematically lower in years prior to the election year, affecting the last tax bill paid by property owners prior to voting. Furthermore, properties are undervalued – compared to sales data – by approximately 1 percent in the year preceding a potential election.

Moreover, leveraging local-election data, I find that counties in which an incumbent appraiser eventually faces a challenger have property values that are undervalued to a greater degree – an additional 2.6 percent. This occurs during the year preceding the contested election – the one that affects the last tax bill paid prior to voting. These results suggest that appraisers may not consistently match appraised values to fair-market values when facing political competition.

Background & Literature

In 2023, appraisers were elected in 18 states, appointed in 16 states, and 16 states used a mixed system for selection (Lincoln Institute of Land Policy, 2025). This current state reflects a shift in philosophy over time towards a professionalization of assessment – and a move towards the appointment of assessors by other officeholders rather than election. However, the modal method of selection is still election.

In Florida, counties elect a local property appraiser every four years. All counties except one have elections during presidential election years.[2] Elections may not be held if only one individual files for candidacy. Of the three election cycles in the time frame (2013-2024), a potential 198 elections could have been held yet only 86 elections were conducted. Of these, most were among individuals new to the position – 24 involved challenged incumbents. Of the 24 challenged incumbents, nearly all successfully achieved reelection.

Understanding the sequence of events regarding appraisal helps with interpretation of the paper's results. The process begins on January 1, the date of assessment. Property appraisers are to determine the fair-market value as of that date. Assessments are due to local taxing authorities on June 1; notices of just values are mailed to property owners before August 25; yet the final tax bill is not mailed until November. Payments are not due until March 31 of the following year.[3]

If a property appraiser desired to influence voter behavior, suppressing values *the year prior* to the election year would ensure the chance to affect the last tax bill paid prior to an election. Suppressing values during the election year would not necessarily engender support given primary elections held in mid-August, right around the deadline to notify property owners but before official tax bills are mailed out. On the other hand, the qualifying deadline for candidates of local elections is during the election year, so appraisers – in the year prior – would not definitely know that their seat would be challenged.

In Florida, all appraisers are elected to four-year terms per the state constitution. The appraiser's primary objective is to provide the county with a fair-market value of all property within the county each year. This is referred to as the “just value” in Florida. Using the just value, one may then derive the “assessed value” and “taxable value”, the last of which is used in conjunction with a jurisdiction's millage rates to determine tax liability.[4] However, the most important figure to consider is the one over which the appraiser has control, the just value.

Researchers have considered the tradeoffs between appointing and electing assessors. Localities with appointed assessors are more likely to achieve assessment quality (Kim et al., 2020) and reassess property values (Eom et al., 2017), and appraised property values grow more quickly under appointed assessors (Makowsky and Sanders, 2013). Elected assessors may undervalue property relative to appointed assessors (Ross, 2011; Ross, 2013; though see Bowman and Mikesell, 1989). Even given these limitations, elected assessors may be more cost efficient than appointed ones (Santerre and Bates, 1996; Propheter, 2015).

While these studies examine and discuss the incentives faced by elected appraisers, most are a wholesale comparison of the two groups. This paper narrows the focus to studying behavior within the institutional framework of elected assessors only, comparing election season to non-election years. I do not attempt to compare elected to appointed officials, rather to isolate and identify the behavior of officials during times of particularly high political incentives (e.g. a challenged incumbency). Given the importance of the tax, it is crucial to understand if the politicization of the office has deleterious or beneficial effects for local governments.

Data

I use county-level property value data from Florida spanning 2013-2024, provided by the Florida Department of Revenue. The data are county-level aggregates. I use real values throughout the analysis. The department also provides a measure of accuracy for appraisers: the just price to sales price ratio. This matches the appraised value on January 1 of a given year to all the property sold during that year. A value of 1 implies the property assessor in county c and year y perfectly appraised property that eventually was eventually sold in that county during that year. Values less than 1 imply that the assessor undervalued property; values greater than 1 imply that the assessor overvalued property.

See Table 1 for summary statistics of variables in the analysis, reported at the county-year unit of observation. On average, the total value of a county’s property in Florida was $46.4 billion. Also, the just price to sales price ratio was 1.005, implying that assessors on average accurately valued property during the timeframe. Note that the official values of the just price to sales price ratio have excessive minimum and maximum values; however, see Figure 1 for the variable’s distribution, which shows that the maximum and minimum are true outliers. To insure against potential errors in the data, I exclude the fifth and tenth percentiles on each end of the distribution as a robustness check in the Model and Results section (e.g. the minimum and maximum – after lopping off the tenth percentile on each end of the distribution – are 0.934 and 1.107). Results remain unchanged.

Election data come from the Florida Division of Elections, within the Department of State. I link primary and general election data from 2012-2024 to identify which counties have incumbent races in a given election cycle.

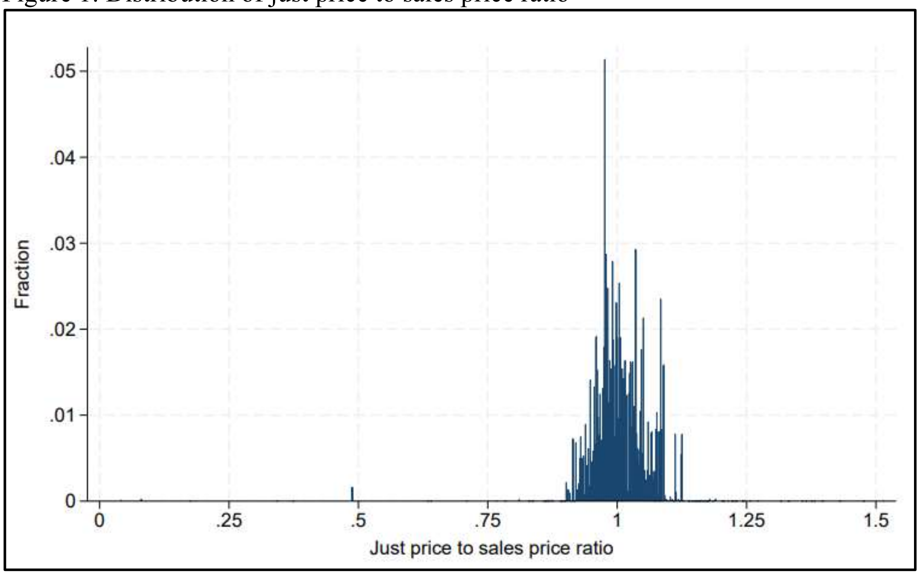
Aside from the assessment data, I also use control variables from the Bureau of Economic Analysis and the Bureau of Labor Statistics.

Table 1: Summary statistics

Variable	Mean	SD	Minimum	Median	Maximum
Just value ($ billions)	46.4	74.7	0.9	19.8	514.1
Just price to sales price ratio	1.005	0.54	0.040	1.000	1.515
Population (thousands)	280.5	398.1	7.3	119.0	1,962.5
Per capita personal income ($ thousands)	53.3	20.0	23.2	48.4	137.9
Unemployment rate	4.9	1.8	1.9	4.5	14.1
Total number of parcels (thousands)	143.9	164.5	5.7	92.8	755.6

Note: All data are for county c in year y from 2013-2024 in Florida. Miami-Dade County is excluded. Just price to sales price ratio is weighted by sales price. Valuation and parcel data come from the Florida Department of Revenue. Population and personal income data come from the Bureau of Economic Analysis. The unemployment rate is from the Bureau of Labor Statistics. While the just price to sales price ratio has notably extreme minimum and maximum values, see Figure 1. Number of observations: 792.

Figure 1: Distribution of just price to sales price ratio



Note: Data come from the Florida Department of Revenue and are weighted by sales price.

Model and Results

All Election Years

I use a two-way fixed-effects model to identify any statistically different property values in election years compared to those without. If assessors are systematically lowering values in order to curry favor with voters, one might expect to see lower than usual values during election season.

I estimate

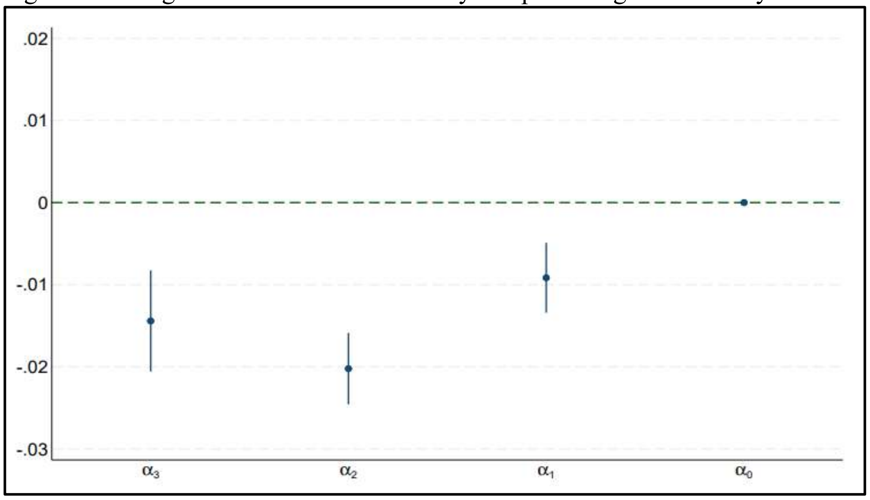
$$AppraisedValue_{cy} = \sum_{i=1}^{3} \alpha_i RelativeYear_{cy}^{i} + \boldsymbol{X}_{cy}\boldsymbol{\theta} + \tau_y + \eta_c + \epsilon_{cy} \tag{1}$$

where *AppraisedValue* is the logged just value all property of county c in year y; $RelativeYear^i$ is equal to 1 for year *t-I* relative to election-year t; $\boldsymbol{X}$ includes county-level control variables (logged population, logged per capita income, logged unemployment rate, logged number of parcels); and τ_y and η_c represent year and county fixed effects. Given election-year t, α_i estimates the (percent) difference in county-level property values in year *t-i* relative to the election year.

Importantly, the estimates of α_i are for all counties, whether or not an election was held. They identify the change in property values *given the threat of facing an election* in year t.

Figure 2 and Table 2 present results from Equation (1) and show that property values are on average increasing as the election year approaches – assessed values are actually highest during election years. However, in year *t-1*, which is the basis for the last property tax bill a property owner pays, assessed values are roughly 1 percent lower than election years.

Figure 2: Change in assessment values for years preceding an election year



Note: Estimates from Equation (1). α_0 represents an election year. The dependent variable is logged appraised (just) value. Data span 2013-2024. Number of observations: 792. Confidence intervals are at the 95 percent level. See Table 2 for more details.

Table 2: Change in assessment values for years preceding an election year

Dependent variable: Logged appraised value	
RelativeYear^3	-0.014***
	(0.003)
RelativeYear^2	-0.020***
	(0.002)
RelativeYear^1	-0.009***
	(0.002)

Note: Reported are estimates from Equation (1) with the dependent variable equal to the logged appraised (just) value. The variable *RelativeYear*^{i} is equal to one for year *t-i* relative to election year *t*. The reference category is an election year. Other independent variables are logged population, logged real per capita income, logged unemployment rate, and logged number of parcels. County and year fixed effects are included. Robust standard errors are in parentheses. Data span 2013–2024. Number of observations: 792. * p<0.10, ** p<0.05, *** p<0.01

While the results suggest that appraised property values are at their highest during election years, that could be the case simply because *actual* (i.e. fair-market) property values are highest then, too. Then appraisers would simply be accurately representing market value. To observe how accurate assessors are, one may take their fair-market estimates given at the beginning of the year and compare their estimate with the actual sale price of property sold during the year. If the values are close, then the assessor correctly approximated the just value. If not, then the appraiser over- or underestimated the value of the property relative to true market values. Observing these values over time can provide a measure of consistency.

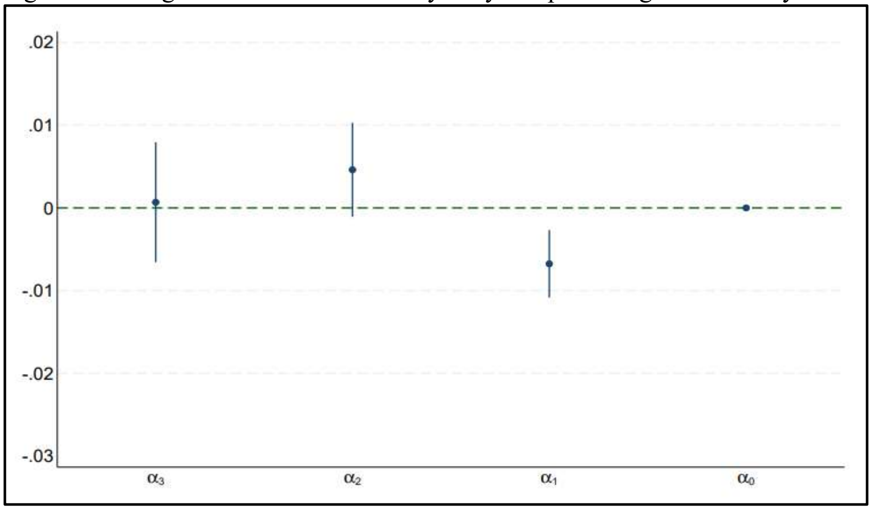
Figure 3 and Table 3 illustrate the results from an estimation of Equation (1) with the ratio of the just valuation to sales price as the dependent variable. This is calculated using all sold property within a county during a given year. A value of one for the dependent variable implies that the assessor exactly matched the future sales price. A value less than one implies that the assessor undervalued the property, and a value greater than one implies that the assessor overvalued the property.

The estimate of α_1 is negative and statistically different from zero. This implies that assessors – in the year prior to an election year – on average underestimate (re fair-market value) the value of property within their county by 1 percent, relative to estimates during all other years.[5]

Challenged Incumbents

Now that the general pattern of assessments relative to election years has been established, do counties in which assessors ultimately face challengers exhibit differential outcomes? To address this question, I examine the case of challenged incumbents.[6]

Figure 3: Change in assessment accuracy for years preceding an election year



Note: Estimates from Equation (1) using the ratio of just value to sales price as the dependent variable. The regression is weighted by total sales price. α_0 represents an election year. Data span 2013-2024. Number of observations: 792. Confidence intervals are at the 95 percent level. See Table 3 for more details.

Table 3: Change in assessment accuracy for years preceding an election year

Dependent variable: Just value to sales price ratio	
RelativeYear^3	0.001
	(0.004)
RelativeYear^2	-0.005
	(0.003)
RelativeYear^1	-0.007***
	(0.002)

Note: Reported are estimates from Equation (1) with the dependent variable equal to the ratio of just value to sales price. The regression is weighted by total sales price. The variable *RelativeYear*^{i} is equal to one for year *t-i* relative to election year *t*. The reference category is an election year. Other independent variables are logged population, logged real per capita income, logged unemployment rate, and logged number of parcels. County and year fixed effects are included. Robust standard errors are in parentheses. Data span 2013-2024. Number of observations: 792. * p<0.10, ** p<0.05, *** p<0.01

In general, one may expect publicly-elected appraisers who are facing a challenger during their re-election campaign to be the most likely to suppress assessed property values in order to elicit favor from voters. To test this hypothesis, I estimate

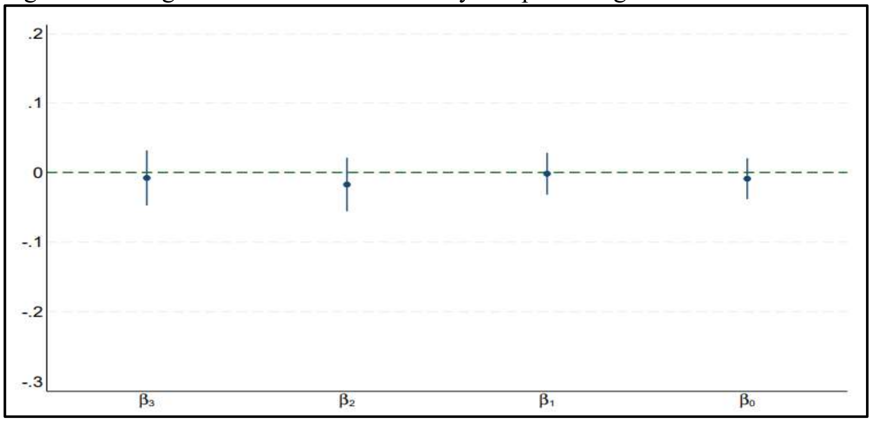
$$AppraisedValue_{cy} = \sum_{i=1}^{3} \alpha_i RelativeYear_{cy}^{i} + \sum_{j=0}^{3} \beta_j Challenged_{cy}^{j} + \boldsymbol{X}_{cy}\boldsymbol{\theta} + \tau_y + \eta_c + \epsilon_{cy} \tag{2}$$

where *Challenged*^{j} is equal to 1 for year *t-j* relative to election-year *t* in which the incumbent is challenged. β_j estimates the (percent) difference in county-level property values in year *t-j* for counties with a challenged incumbent relative to counties without a challenged incumbent. Essentially, β_j is the *differential* change in assessed just values by a challenged incumbent – over and above any effects applying to all assessors, still captured by α_i.

Figure 4 and Table 4 show the differential *challenged-incumbent* effect. The fact that the estimates of β_j are not distinguishable from zero implies that the appraised just value of property is no different in counties with challenged incumbents than in those without such. Moreover, the estimates of α_i from Equation (2) are remarkably similar to those from the estimation of Equation (1), implying that the pattern depicted in Figure 2 apply, on average, to all counties.

In a similar manner to the analysis depicted in Figure 3, I estimate Equation (2) with the ratio of the just valuation to sales price as the dependent variable. Results are displayed in Figure 5 and Table 5. The coefficient β_1 is negative, statistically significant, and represents the systemic underestimate of property values in the year before a contested election – the last valuation that affects tax liability prior to the election. The estimate of α_1 from Equation (2) is similar to that in Table 3. This result, combined with the estimate of β_1, suggests that, while all counties have undervalued properties in years prior to an election year, those with challenged incumbents experience a greater degree of undervaluation.

Figure 4: Change in assessment values for years preceding an election with a challenged incumbent



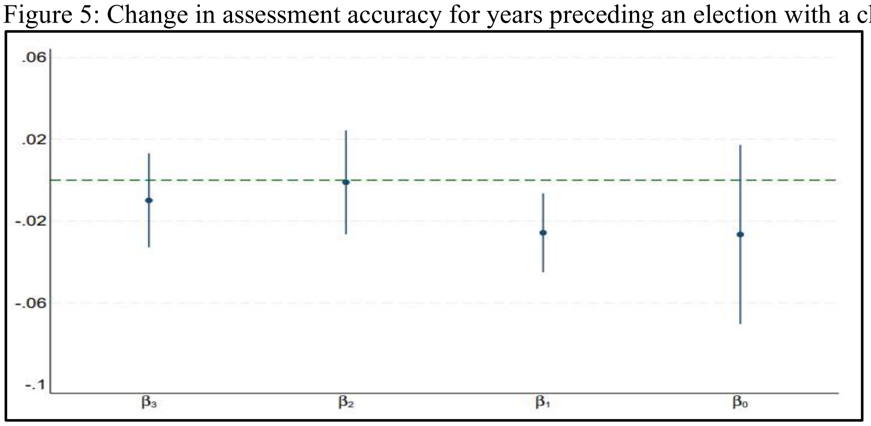
Note: Estimates from Equation (2). β_0 represents an election year. Data span 2013-2024. Number of observations: 792. Confidence intervals are at the 95 percent level. See Table 4 for more details.

Table 4: Change in assessment values for years preceding an election with a challenged incumbent

Dependent variable: Logged appraised value	
RelativeYear^3	-0.015***
	(0.004)
RelativeYear^2	-0.018***
	(0.004)
RelativeYear^1	-0.010**
	(0.004)
Challenged^3	-0.008
	(0.020)
Challenged^2	-0.017
	(0.020)
Challenged^1	-0.002
	(0.015)
Challenged^0	-0.009
	(0.015)

Note: Reported are estimates from Equation (2) with the dependent variable equal to the logged appraised (just) value. The variable *RelativeYear^i* is equal to one for year *t-i* relative to election year *t*. The variable *Challenged^j* is equal to one for year *t-j* relative to election year *t* in which an incumbent is challenged in county *c*. The reference category is an election year. Other independent variables are logged population, logged real per capita income, logged unemployment rate, and logged number of parcels. County and year fixed effects are included. Robust standard errors are in parentheses. Data span 2013–2024. Number of observations: 792. * p<0.10, ** p<0.05, *** p<0.01

Figure 5: Change in assessment accuracy for years preceding an election with a challenged incumbent



Note: Estimates from Equation (2) using the ratio of just value to sales price as the dependent variable. The regression is weighted by total sales price. β_0 represents an election year. Data span 2013-2024. Number of observations: 792. Confidence intervals are at the 95 percent level. See Table 5 for more details.

Table 5: Change in assessment values for years preceding an election with a challenged incumbent

Dependent variable: Logged appraised value	(1) Full sample	(2) 5-95 percentile	(3) 10-90 percentile
RelativeYear^3	-0.003	-0.002	-0.003
	(0.004)	(0.004)	(0.004)
RelativeYear^2	-0.003	-0.000	-0.003
	(0.005)	(0.004)	(0.003)
RelativeYear^1	-0.009*	-0.005**	-0.005*
	(0.005)	(0.002)	(0.002)
Challenged^3	-0.010	-0.004	-0.004
	(0.012)	(0.012)	(0.013)
Challenged^2	-0.001	0.005	0.010
	(0.013)	(0.012)	(0.011)
Challenged^1	-0.026**	-0.018**	-0.010
	(0.010)	(0.008)	(0.009)
Challenged^0	-0.027	-0.006	-0.005
	(0.022)	(0.005)	(0.006)
Total effect for challenged incumbents			
RelativeYear^3 + *Challenged*^{3}	-0.013	-0.006	-0.007
	(0.011)	(0.010)	(0.011)
RelativeYear^2 + *Challenged*^{2}	-0.004	0.005	0.007
	(0.012)	(0.009)	(0.010)
RelativeYear^1 + *Challenged*^{1}	-0.034***	-0.023***	-0.015*
	(0.012)	(0.007)	(0.008)
Challenged^0	-0.027	-0.006	-0.005
	(0.022)	(0.005)	(0.006)

Note: Reported are estimates from Equation (2) with the dependent variable equal to the ratio of just value to sales price. The regression is weighted by total sales price. The variable *RelativeYear*^{i} is equal to one for year *t-i* relative to election year *t*. The variable *Challenged*^{j} is equal to one for year *t-j* relative to election year *t* in which an incumbent is challenged in county *c*. The reference category is an election year. Other independent variables are logged population, logged real per capita income, logged unemployment rate, and logged number of parcels. County and year fixed effects are included. Robust standard errors are in parentheses. Data span 2013-2024. Number of observations: 792 for Column (1); 712 for Column (2); and 632 for Column (3). * p<0.10, ** p<0.05, *** p<0.01

The sum of α_1 and β_1 represent the total effect of decreased accuracy by challenged incumbents – the coefficient is -0.034 (s.e. 0.012), implying that, in the year prior to facing reelection and only during that year, assessors on average undervalue the property in their county by 3.4 percentage points relative to other years. Even more conservative estimates of the total effect, using smaller ranges of data, stand at a reduction of 2.3 and 1.5 percentage points.

Discussion and Conclusion

Over one-third of states use elections to select local property appraisers, and an additional third of states use a mixture of elections and political appointment. Examining data from Florida, a state with local elections, this paper puts forth three key findings: i) on average, elected appraisers appear to assess property at lower values during years that precede a potential election, all else equal; ii) on average, elected appraisers appear to undervalue property (relative to actual sales data) for the final tax bill paid prior to a potential election; and iii) for counties that have challenged incumbent assessors, appraised property values appear to be even more undervalued than in counties that do not have challenged incumbents.

These findings illustrate a potential risk of using elections to choose property appraisers – an erosion of the tax base for one of the most important taxes to local governments. This of course should not be the only consideration when comparing the two primary ways of choosing assessors, but isolating appraisers from direct political forces may increase consistency and accuracy.

On the other hand, the extent of the problem, as estimated, is limited. Of the 66 counties in the sample, over three election cycles, only 24 incumbents were challenged – 12 percent of the election opportunities in the sample. All other races involved completely new participants or the incumbent ran unchallenged. Furthermore, the appraised just value is not the true tax base in Florida, so reductions in the just value lead to less-than-proportionate reductions in the tax base. Finally, given a decrease in just values, taxing authorities can increase millage rates to equalize revenue from year to year.

Looking forward, it would be interesting to see if these findings manifest in other settings. Do other states that elect appraisers exhibit the same patterns documented here? Also, due to limited data availability, I am not able to compare sales prices of single-family homes to appraisal values of single-family homes – the data presented here are for all properties in a county. If the underestimation of value is truly a political artifact, one would perhaps observe stronger effects for single-family homes than other types of property, given voters' incentives. More research is required for a more complete understanding of the political forces at play in this consequential local position.

Acknowledgements

I thank participants at the 2024 Academy of Economic and Finance conference, the 2024 Eastern Economics Association annual meeting, and College of Charleston economics workshop for helpful comments. All errors and omissions are my own.

Notes

1. I use appraiser and assessor interchangeably throughout the paper. Also, this paper is only concerned with the public office of tax appraiser, which is distinctively different from private appraisers who work in conjunction with banks etc.
2. Duval County holds elections in the year prior to a presidential election. Miami-Dade has a unique local government structure that was amended in 2018. Data from Miami-Dade County is omitted entirely from this analysis.
3. See floridarevenue.com/property for full details.
4. To avoid confusion, I use "appraised value" or "just value" to refer to the property appraiser's fair-market valuation of a property and eschew the use of "assessed value".
5. For better interpretation of results, I report percent change. Technically, the unit of the coefficient is percentage points, but with a mean of 1, a 1 percentage point decrease is a 1 percent decrease. I also report percent change when referencing Table 5.
6. I define "challenged incumbents" as those who faced a challenger in either the primary or general election of a given election year. I ignore races for open seats, since the candidates have no direct link to property assessment. There are 24 challenged incumbents in the sample. See the discussion in the Background section.

References

Bowman JH, Mikesell JL (1989) Assessment uniformity under elected versus appointed assessors. National Tax Journal 42:181–189

Eom TH, Bae H, Kim S (2017) Moving beyond the influence of neighbors: Local influences on property tax reassessment decisions in New York. American Review of Public Administration 47:599–614

Kim S, Chung IH, Eom TH (2020) Institutional differences and local government performance: Evidence from property tax assessment quality. Public Performance & Management Review 43:388–413

Lincoln Institute of Land Policy (2025) Assessment administration dataset. Lincoln Institute of Land Policy

Makowsky M, Sanders S (2013) Political costs and fiscal benefits: The political economy of residential property value assessment under Proposition 2 1/2. Economics Letters 120:359–363

Propheter G (2015) Elected and appointed assessor efficiency differentials. Public Performance & Management Review 39:358–380

Ross JM (2011) Assessor incentives and property assessment. Southern Economic Journal 77:776–794

Ross JM (2013) A socioeconomic analysis of property assessment uniformity: Empirical evidence on the role of policy. Public Budgeting & Finance 33:49–75

Santerre RE, Bates LJ (1996) Performance and pay in the public sector: The case of the local tax assessor. Public Finance Quarterly 24:481–493

PARKER
COLLEGE OF BUSINESS
GEORGIA SOUTHERN UNIVERSITY

Department of Economics
Department of Finance
11935 Abercorn Street
Savannah, GA 31419