ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

A MASTER THESIS ON

Do Commodities Improve Traditional Equity-Bond Asset Allocations?

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science (MSc) in Banking & Finance

by

Pantatosakis Pavlos
Student ID: MXRH2423

Supervisor:
Skoulakis Georgios, Associate Professor

Examination Committee:
Anthropelos Michail, Associate Professor
Englezos Nikolaos, Assistant Professor
Skoulakis Georgios, Associate Professor

Piraeus, Greece
March 2026

Acknowledgments

I would like to thank my family and friends for their support and understanding throughout my Master's journey. I would also like to express my gratitude to my professor and supervisor, Georgios Skoulakis, for his guidance, support, and feedback during the course of this thesis. His expertise and advice were essential to the completion of this research.

Abstract

This thesis investigates whether the inclusion of commodities can improve the performance of traditional equity-bond portfolios in an environment of shifting macroeconomic regimes, using data from 1980 to 2021. While the 60/40 stock-bond allocation has historically provided diversification, the tendency of stock-bond correlation to increase during periods of inflation uncertainty and market stress suggests a need for alternative asset classes. We create a dynamic, out-of-sample investment strategy using Time-Weighted Least Squares (TWLS) and an Exponentially Weighted Moving Covariance (EWMC) matrix. This approach allows portfolio weights to adapt to potential structural breaks by prioritizing recent data. Four commonly used macroeconomic predictors are employed to estimate conditional expected excess returns, subject to theoretically motivated sign constraints on the slope coefficients. We evaluate three generations of commodity indices to determine the impact of index construction on portfolio efficiency. The results indicate that while first-generation indices provide limited value, third-generation indices offer diversification benefits, which stay robust after various tests. Specifically, the Certainty Equivalent Return (CER) test confirms that the addition of commodities to a stock-bond portfolio provides statistically significant (p-value $< 10\%$) benefits for risk-averse investors. Ultimately, this study constructs a practical allocation strategy in real time and demonstrates that commodities serve as a tool for enhancing portfolio resilience and capital preservation.

Keywords: Asset Allocation, Portfolio Choice, Commodities

Περίληψη

Η παρούσα διπλωματική εργασία διερευνά κατά πόσο η συμπερίληψη των εμπορευμάτων (commodities) μπορεί να βελτιώσει την απόδοση των παραδοσιακών χαρτοφυλακίων μετοχών-ομολόγων σε ένα περιβάλλον μεταβαλλόμενων μακροοικονομικών συνθηκών, χρησιμοποιώντας δεδομένα για την περίοδο 1980-2021. Ενώ μια παραδοσιακή κατανομή 60% σε μετοχές και 40% σε ομόλογα ιστορικά δείχνει να προσφέρει οφέλη από την διαφοροποίηση, η τάση της συσχέτισης μεταξύ μετοχών και ομολόγων να αυξάνεται σε περιόδους πληθωριστικών πιέσεων και αβεβαιότητας στις αγορές, καθιστά αναγκαία την αναζήτηση εναλλακτικών επενδύσεων. Ακολουθούμε μια δυναμική στρατηγική, χρησιμοποιώντας τη μέθοδο των χρονικά σταθμισμένων ελαχίστων τετραγώνων και έναν εκθετικά σταθμισμένο πίνακα συνδιακύμανσης με χρονικά μεταβαλλόμενο μέσο. Η προσέγγιση αυτή επιτρέπει στα βάρη του χαρτοφυλακίου να προσαρμόζονται σε ενδεχόμενες δομικές αλλαγές στην οικονομία, δίνοντας περισσότερη βαρύτητα στα πιο πρόσφατα δεδομένα. Για την πρόβλεψη των αναμενόμενων πριμ, χρησιμοποιούνται τέσσερις διαδεδομένοι μακροοικονομικοί δείκτες, στους οποίους επιβάλλουμε περιορισμούς στο πρόσημο τους, βάσει της θεωρίας. Αξιολογούμε τρεις γενιές δεικτών εμπορευμάτων για να προσδιορίσουμε την επίδραση της δομής του κάθε δείκτη στην αποτελεσματικότητα του χαρτοφυλακίου, και τα αποτελέσματα δείχνουν ότι ενώ οι δείκτες πρώτης γενιάς προσφέρουν περιορισμένη αξία, οι δείκτες τρίτης γενιάς παρέχουν σημαντικά οφέλη διαφοροποίησης. Συγκεκριμένα, ο έλεγχος για το ισοδύναμο βεβαιότητας (Certainty Equivalent Return - CER) επιβεβαιώνει ότι η προσθήκη εμπορευμάτων σε ένα χαρτοφυλάκιο μετοχών-ομολόγων προσφέρει στατιστικά σημαντικά οφέλη (p-value $<$ 10%) για επενδυτές που αποστρέφονται τον κίνδυνο. Στη

μελέτη αυτή δεν αναλύουμε τα θεωρητικά οφέλη των εμπορευμάτων, αλλά κατασκευάζουμε μια στρατηγική κατανομής βαρών σε πραγματικό χρόνο. Τελικά, η διατριβή αυτή καταδεικνύει ότι τα εμπορεύματα αποτελούν ένα εργαλείο για την ενίσχυση της ανθεκτικότητας του χαρτοφυλακίου και τη διατήρηση του κεφαλαίου.

Λέξεις-Κλειδιά: Κατανομή Περιουσιακών Στοιχείων, Επιλογή Χαρτοφυλακίου, Εμπορεύματα

Contents

1 Introduction

In recent years, the global economy has experienced significant price volatility due to geopolitical tensions, inflation, and shifts in economic policy. These challenges could lead investors to reconsider traditional asset allocation strategies, such as the 60/40 stock-bond portfolio, which may be sub-optimal, as other asset classes that could provide diversification benefits are omitted. Commodities may help investors and portfolio managers during such periods, as they could contribute to portfolio resilience and act as an inflation hedge (Bampinas and Panagiotidis, 2015). For instance, Gold often retains or increases in value when the purchasing power of fiat currencies declines, as the drivers of commodity returns are usually different from those of other asset classes (Baur and McDermott, 2010). Moreover, commodities have risk and return characteristics that differ from those of equities and fixed-income, and this could make them a powerful tool for diversification, especially when traditional asset classes underperform.

The purpose of diversification is to mitigate the impact of events that reduce portfolio returns by ensuring that gains from some investments offset losses from others. This effect arises when asset classes are negatively correlated or have relatively low correlation. Over the past two decades (2000–2020), the average correlation between stocks and bonds has been negative, meaning that these two asset classes generally moved in opposite directions. As a result, market participants and investors are unfamiliar with a positive stock-bond correlation, even if this is not unusual; since 1875, periods of positive correlation have been more frequent than periods of negative correlation (Molenaar et al., 2024). Various studies show that during crises and uncertainty, alternative asset classes could positively impact overall portfolio performance. Bekkers et al. (2009) suggest that adding alternative asset classes (e.g., real estate & commodities) to a traditional asset class portfolio (equity & fixed-income) can be the most efficient allocation, adding value for investors. Bodie and Rosansky (1980) found that a portfolio consisting of 60% equities and 40% commodity futures reduces volatility by one third without lowering overall returns.

Between 1998 and 2007, the volume of U.S. exchange-traded futures and futures options expanded nearly fivefold, reaching 3.2 billion contracts in 2007 (Staff Report on Commodity Swap Dealers & Index Traders with Commission Recommendations, 2008). This growth reflects the increasing participation of financial investors, who have contributed to the financialization of commodity markets through trading (Domanski and Heath, 2007). Although such participation can enhance market efficiency and heighten liquidity, the financialization effect brings risks similar to those in other segments of the financial system. An important driver of this process was the entry of hedge funds and other speculative participants into the commodity futures markets during the 2000s (Büyükşahin and Robe, 2014). Basak and Pavlova (2016) show that the increasing integration of financial investors into commodity markets raises concerns regarding the prices of commodities and the spillovers it can bring to the economy, but also can offset the traditional diversification role of commodities. Silvennoinen and Thorp (2013) find that commodity–equity correlations were close to zero during the 1990s and climbed to approximately +0.5 at the peak, showing that diversification benefits to investors across equity, bond and stock markets were significantly reduced.

The purpose of this study is to provide insights into the benefits and risk reduction that commodities can offer to a stock-bond allocation, and our objective is to identify if the inclusion of commodities may outperform the traditional portfolio.

2 Literature Review

The regime that existed from 1970 to 2000 was characterized by a positive correlation between equities and bonds, and the reappearance of this relationship could increase portfolio variance and raise investment risk. Brixton et al. (2023) find that changes in fundamental macroeconomic factors, particularly inflation uncertainty, may lead to a positive correlation between these two asset classes. Their analysis considers both industrial growth and inflation volatility and they find that during periods of high growth uncertainty, the stock-bond correlation tends to remain negative; in contrast, when inflation uncertainty increases, the correlation is more likely to turn positive. Although different events drive the returns of the three aforementioned asset classes, they share common factors due to globalization, international trade and the financialization of commodities that have increased correlations among them. Commodities, however, are a broad asset class and should be considered in subgroups. Štulec et al. (2024) analyze whether all commodities are fundamentally similar and suggest that Gold is a suitable choice for risk-averse investors, whereas crude Oil may be preferable for risk-tolerant investors. Additionally, commodities traded in the spot market, either directly or via exchange-traded funds (ETFs), can benefit long-term investors, while commodity futures may be more efficient for short-term strategies (e.g., during crises or market correction) (Anson, 1998).

Historical returns are often poor predictors of future performance, meaning that in-sample results tend to overestimate the diversification benefits of commodity investments (Bessler and Wolff, 2015). Therefore, out-of-sample analysis offers a more realistic approach, as it accounts for the estimation risk surrounding expected returns. Ruano and Barros (2022) employed different allocation strategies (strategic weighting, risk parity, and mean–variance) and accounted for behavioral aspects such as loss aversion, rather than assuming purely rational, risk-averse investors. Their results indicate that while commodity futures continue to provide diversification, the benefits are considerably weaker in out-of-sample tests than those suggested by in-sample analysis.

Hull (2014) (Chapter 5.14) discusses how futures prices relate to spot prices and the risks faced by investors who take long or short positions in futures contracts. He then examines the theoretical link between the current futures price ($F_{0,T}$) and the expected spot price at maturity $\mathbb{E}[S_T]$, showing that the futures price can be seen as an unbiased estimate of the future spot price when the asset's returns are uncorrelated with overall market returns. If the futures price is lower than the expected spot price, the market is referred to as being in normal backwardation and when it is higher, the market is in contango (Figure 1). In the same Chapter, he explains that when hedgers take short positions and speculators long positions, the futures price tends to be below the expected future spot price (normal backwardation). Conversely, when hedgers are in long positions and speculators take short positions, the futures price will be above the expected future spot price (contango). Since the expected spot price at maturity cannot be observed directly, this is a theoretical perspective and in practice, the terms backwardation and contango are usually used to describe the relationship between current spot (S_0) and futures prices ($F_{0,T}$) rather than futures prices and expected future spot prices.

As futures contracts expire, investors need to roll their positions to maintain exposure to commodities. This process involves closing out an existing expiring contract and opening a new one with a later expiration date, with the (positive or negative) return generated from this process referred to as the roll yield (Figure 2).

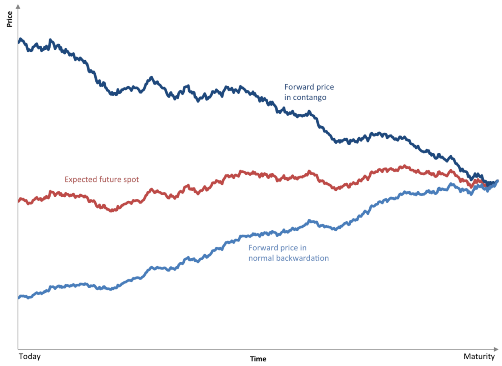

Figure 1: Contango versus backwardation.

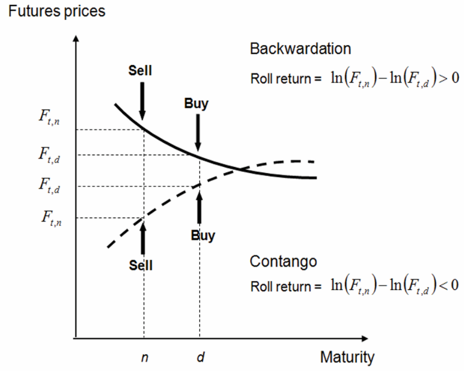

Figure 2: Roll Returns during backwardation & contango, from Miffre (2012).

Taking long positions in backwardated markets can be advantageous, as investors benefit from a positive roll yield, rolling over expiring contracts into cheaper futures. Conversely, in contango, the roll yield is negative as investors must purchase more expensive futures, which reduces their overall returns. This notion aligns with Dunsby and Nelson (2010) who find that the Sharpe Ratio (henceforth SR) of Oil futures in backwardation periods is 0.85, but when the basis is negative (contango), the average monthly return is -0.31%, providing a negative SR during 1991-2010.

In the last three decades, the growth of commodity investments and the introduction of new financial instruments have driven the evolution of commodity indices through three distinct generations, marking a new era in the way investors access the commodity markets.

First-generation commodity indices

Examples: S&P Goldman Sachs Commodity Index (S&P GSCI); Dow Jones-UBS Commodity Index (DJ-UBSCI)

These indices are passive and aim to provide broad exposure to the commodity market by tracking a wide basket of commodities. Their primary goal is market coverage, rather than attempting to outperform through active weighting or risk management strategies. They are production-weighted, which results in a heavy concentration on the energy sector and provide limited benefits to investors seeking a well-diversified commodity portfolio. Furthermore, they are long-only and roll their positions into the next contract, which leads to negative roll yields during steep contango periods, reducing their overall returns.

Second-generation commodity indices

Examples: S&P GSCI Equal Weight Select; UBS Bloomberg Contant Maturity Commodity Index (CMCI); Barclays Commodity Curve Allocation Index

They mitigate the negative effects of first-generation indices during contango periods by investing across the whole maturity curve. Second-generation indices are more diversified, as they provide different weighting strategies (e.g., equal weight, liquidity, sector-capped etc) and will outperform (underperform) the first-generation indices during contango (backwardation) periods. Nevertheless, they are less liquid than first-generation indices, carry higher costs, and share the long-only structure of first-generation indices, resulting in negative roll yields during contango periods.

Third-generation commodity indices

Examples: Morningstar Long/Short (LS) Commodity Index; BNP Paribas COMAC LS Commodity Index

Third-generation indices build upon the limitations of first and second generation, by adopting active, long/short strategies that aim for higher returns and reduced volatility. These indices seek to capitalize on backwardation by taking long positions, while profiting from contango by taking short positions. They mitigate or even reverse the negative roll yield that affects the other two generations; hence, this allows them to generate returns from both rising and falling commodity prices. In addition, the use of non-traditional weighting schemes is common (i.e., overweight (underweight) high (low) expected return commodities) and are typically well diversified, allowing for a more efficient portfolio construction. However, their active nature carries higher trading costs and greater complexity compared to passive commodity indices.

Miffre (2012) analyzes the three generations of commodity indices (Table 1) and evaluates their performance in the period following the Great Recession. She finds that between 2008-2012, first-generation indices recorded negative average excess returns, with a SR of -0.23, whereas second-generation indices had an average SR close to zero. In contrast, third-generation indices achieved an average annualized excess return of 3% with an SR of 0.26, demonstrating a superior performance than the other two generations during the crisis.

This performance gap is further supported by empirical findings which suggest that first-generation indices do not provide benefits to investors, while the evidence for second-generation indices remain mixed, particularly due to their long-only exposure and limited out-of-sample improvement in portfolio efficiency. In contrast, third-generation indices can increase returns and decrease volatility, improving the efficient frontier (Kremer, 2015).

Many studies fail to demonstrate out-of-sample benefits from commodity investments, a limitation that can be linked to both methodological choices and the data used. Research applying mean-variance spanning tests often reports in-sample advantages; however, these benefits typically weaken considerably when tested for out-of-sample robustness. Furthermore, much of the literature relies on first-generation commodity indices, such as the S&P Goldman Sachs Commodity Index (SPGSCI) or the Dow Jones UBS Commodity Index (DJUBS), which may constrain the generalizability of the results due to their suboptimal construction (Rallis et al., 2013). Yan and Garcia (2017) find no benefits from first-generation indices and suggest that investors are likely to shift toward third-generation indices. Similarly, Fethke and Prokopczuk (2018) shows that first-generation indices are spanned by stocks and bonds over the full sample, whereas third-generation indices demonstrate robust out-of-sample benefits.

Daskalaki and Skiadopoulos (2011) investigate the diversification benefits of commodities using both mean-variance and non-mean-variance spanning tests. The study finds that while commodities appear to enhance portfolio performance in-sample for non-mean-variance investors (investors considering higher moments), these benefits diminish out-of-sample, suggesting that traditional spanning tests may overstate the diversification potential of commodities. Daskalaki et al. (2017) reevaluate the benefits of commodities by employing a stochastic dominance efficiency (SDE) approach, which captures the full return distribution and is applicable to all risk-averse investors. Their findings indicate that commodities provide significant diversification benefits both in-sample and out-of-sample. Moreover, these benefits are evident when investors access the commodity market through second or third generation indices, which are better diversified than first-generation indices as discussed above.

Table 1: Comparison of Commodity Index Generations (Miffre (2012)).

Generation / Feature	**First-Generation**	**Second-Generation**	**Third-Generation**
Exposure	Long only	Long only	Long and short
Objective	Broad representation of commodity markets	Mitigate negative roll yield impact while maintaining liquidity	Improve returns and reduce risk via long/short dynamic strategies
Weighting	Production-based	Liquidity and roll-based	Signal-driven
Rolling Strategy	Nearest contract	Advanced methods[a]	Dynamic
Rebalancing	Infrequent (annual)	Periodic (monthly/quarterly)	Based on market conditions
Diversification/ Sector Bias	Often energy-heavy	Consider liquidity and term structure; some sector diversification	Diversified; dynamic hedging reduces sector risk

[a] *Enhanced roll, constant maturity, implied roll yield, etc.*

3 Data

For stocks, we use data retrieved from the Amit Goyal Dataset and the primary source of the data is the Center for Research in Security Prices (CRSP's) calculation of the Standard and Poor's 500 (S&P 500) returns including dividends (variable name: ret). Returns are weighted by market capitalization, meaning larger firms have proportionally more influence in the index. From the same dataset, we obtain the U.S. 3-month Treasury bill rate (variable name: tbl) as a proxy for the risk-free rate. (For more information please refer to the: CRSP info guide and Amit Goyal Spreadsheet.)

We use the Bloomberg US Aggregate Bond Index (LBUSTRUU in Bloomberg) as a proxy for the US fixed-income market. This index reflects the performance of investment-grade bonds, including Treasury securities, government agency bonds, corporate bonds, asset-backed securities, and mortgage-backed securities. It serves as a broad benchmark and is widely used in the literature to capture fixed-income market performance.

To represent commodities, we use indices from all three generations, as discussed in the literature review section. For the first-generation, we use the S&P Goldman Sachs Commodity Index (SPGSCITR), which is a widely recognized benchmark for tracking the performance of global commodity markets. The index covers a broad range of commodity futures across key sectors such as energy, metals, agriculture, and livestock. At the same time, it only includes highly liquid contracts that meet strict trading and volume requirements, ensuring tradability. SPGSCITR is production-weighted, meaning that each commodity's weight reflects its relative importance in the global economy, and as a result, the SPGSCI is heavily weighted towards the energy sector. Bloomberg Commodity Index (BCOMTR) is a more diversified alternative for a first-generation index due to its capping rules that limit overweighting (Dunsby and Nelson, 2010). Nevertheless, both indices assume a backwardated market (Kremer, 2015), which can limit the diversification benefits of commodities as an asset class. To address these limitations, we also examine later-generation indices provided by Morningstar[1], which are designed to deliver improved diversification and performance. For the second-generation of commodities, we use the Morningstar Long/Flat (MSDILFER) and for third-generation the Morningstar Long/Short (MSDILSTR). The Long/Short Commodity Index is a fully collateralized commodity futures index that uses the momentum rule to determine which commodity is held long or short. The Morningstar Long/Flat differs from traditional (first-generation) long-only indices by replacing short positions from the Morningstar Long/Short Commodity Index with flat positions, aiming to capture upside while mitigating downside risk.

Table 2: Asset classes, definitions and their sources.

Variable	**Description**	**Source**
CRSP	CRSP US Equity Index (inc. dividends)	Amit Goyal
LBUSTRUU	Bloomberg US Aggregate Bond TRI	Bloomberg
SPGSCITR	S&P Goldman Sachs Commodity TRI	Bloomberg
MSDILFER	Morningstar Long-Flat Commodity TRI	Bloomberg
MSDILSTR	Morningstar Long-Short Commodity TRI	Bloomberg
TBL	Short-Term Treasury Bill Rate	Amit Goyal

For modeling conditional expected premia, we use four standard and widely used macroeconomic variables as return predictors. The Treasury bill rate (TBL), which serves as a proxy for the discount rate; the dividend-to-price ratio (DP), capturing equity market valuation; the term spread (TMS), defined as the difference between

[1] Construction Rules for Morningstar Commodity Indexes.

long-term and short-term government bond yields and reflecting the slope of the yield curve; and the default yield spread (DFY), measured as the BAA minus AAA spread and serving as an indicator of credit risk. All variables are obtained from the Amit Goyal dataset. These predictors have been widely used in the literature on equity premium predictability, with a large number of studies published in leading finance journals (Goyal et al., 2024).

Table 3: Predictors, definitions and their sources.

Variable	**Description**	**Source**
TBL	Short-Term Treasury Bill Rate	Amit Goyal
DP	Log Dividend-to-Price Ratio	Amit Goyal
TMS	Term Spread (Long - Short Term Yield)	Amit Goyal
DFY	Default Yield Spread (BAA-AAA)	Amit Goyal

The Treasury bill rate (TBL) is identified by Ang and Bekaert (2007) as the most robust predictive variable for equity returns at short horizons, where it exhibits a strong negative relationship with stock performance. This finding aligns with evidence from Fama and Schwert (1977) and is confirmed to be robust across multiple international markets, often providing more consistent evidence of predictability than the dividend yield alone. More recent work by Avramov and Chordia (2006) and Rapach et al. (2013) continue to utilize the Treasury bill as a primary business cycle indicator, demonstrating its value for out-of-sample forecasting.

The dividend-price ratio (DP) is a standard equity return predictor in financial literature, reflecting the intuition that stock prices are low relative to dividends when discount rates and expected returns are high. Fama and French (1988) and Campbell and Shiller (1988) demonstrate that its predictive power, measured by regression's R-Squared, increases with the return horizon, explaining a portion of multiyear return variances. Cochrane (2008) argues that the near absence of dividend growth predictability in the data necessitates that dividend yield variation must be driven by changing expected returns. While univariate predictive power may be weaker at short horizons, Ang and Bekaert (2007) note that its ability is considerably enhanced when used in a bivariate regression alongside the short rate (TBL) as discussed above.

The term spread (TMS), calculated as the difference between long-term and short-term government bond yields, is used to identify time-varying maturity premia that are common to both stocks and bonds. Fama and French (1989) and Fama (1990) document that this spread follows a business-cycle pattern, typically being low near economic peaks and high near troughs. Estrella and Hardouvelis (1991) further establishes its significance by showing that the slope of the yield curve can predict cumulative changes in real output and the probability of a recession.

The default yield spread (DFY), defined as the yield differential between lower-grade (BAA) and higher-grade (AAA) bonds, serves as a proxy for business conditions and credit risk. Chen et al. (1986) and Fama and French (1989) argue that this spread captures predictable components of returns that are high during periods of persistent economic weakness (e.g., the Great Depression) and low when business conditions are strong. Studies by Keim and Stambaugh (1986) confirm that the default spread forecasts returns on both bonds and stocks, with its regression slopes increasing for riskier assets.

Using these variables, we aim to capture the performance of stocks, bonds, and commodities, with the risk-free rate used to calculate the excess returns of the three asset classes. Observations are recorded at a monthly frequency, which is commonly used in asset pricing literature. Our dataset spans the time period from January 1980 to May 2021, providing 497 return observations, covering 40 years of data.

3.1 Descriptive statistics

Table 4 presents descriptive statistics and correlations for the five assets from January 1980 to May 2021. Panel A shows that CRSP US Equity generated the highest compound annual growth rate (12.29%), while Bloomberg US Agg (bonds) offered the best risk-adjusted return with a SR of 0.61. In contrast, the S&P Goldman Sachs index (first-generation of commodities) underperformed, showing the highest volatility (20.09%) and a low SR (0.05). Panel B reports the correlation matrix, highlighting a low correlation (+0.17) between equities and bonds in the full sample, which supports the diversification benefits of combining these assets in the traditional portfolio. However, the Morningstar Long/Short index (third-generation of commodities), demonstrates strong risk-adjusted returns, while being uncorrelated with the CRSP and the Bloomberg US Agg fixed-income index. In Table 5, we present the summary statistics of the four predictors.

Table 4: Summary statistics and correlations of all asset classes: 1980–2021.

Panel A: Descriptive Statistics						
	CAGR	Premium	Volatility	SR	Skew	Kurt
CRSP (Equity)	12.29%	8.69%	15.11%	0.58	-0.60	2.07
LBUSTRUU (Fixed Income)	7.37%	3.17%	5.19%	0.61	0.42	6.31
SPGSCITR (First Gen.)	3.16%	1.08%	20.09%	0.05	-0.38	2.71
MSDILFER (Second Gen.)	3.20%	-0.50%	9.60%	-0.05	0.13	3.14
MSDILSTR (Third Gen.)	9.67%	5.69%	10.13%	0.56	0.34	1.90

Panel B: Pearson Correlation						
CRSP (Equity)	1.00	0.17	0.23	0.11	-0.02	0.00
LBUSTRUU (Fixed Income)	0.17	1.00	-0.04	-0.07	-0.01	0.15
SPGSCITR (First Gen.)	0.23	-0.04	1.00	0.68	0.39	0.04
MSDILFER (Second Gen.)	0.11	-0.07	0.68	1.00	0.77	-0.04
MSDILSTR (Third Gen.)	-0.02	-0.01	0.39	0.77	1.00	0.21
Risk-Free Rate	0.00	0.15	0.04	-0.04	0.21	1.00

Note: Panel A reports Compound Annual Growth Rate (CAGR), average annualized excess returns (Premium), annualized volatilities (Volatility), annualized Sharpe ratios (SR), skewness (Skew) and excess kurtosis (Kurt). Panel B reports pairwise correlations between asset classes simple returns. The sample consists of 497 observations: January 1980 - May 2021.

Table 5: Descriptive statistics for all predictors. Treasury Bill (TBL), Term Spread (TMS) and Default Yield Spread (DFY) as percentages (%) and the logarithm of dividend-to-price (DP). The sample consists of 497 observations: January 1980 - May 2021.

Summary	**TBL (%)**	**DP**	**TMS (%)**	**DFY (%)**
Min.	0.00	−4.52	−3.65	0.55
1st Qu.	0.08	−4.03	1.10	0.77
Median	0.33	−3.87	2.24	0.95
Mean	0.34	−3.75	2.14	1.08
3rd Qu.	0.48	−3.44	3.29	1.27
Max.	1.36	−2.75	4.55	3.38

Figure 3 illustrates the time-varying nature of asset class correlations, using a rolling window of 60 months. Consistent with the literature (Molenaar et al., 2024), the graph demonstrates that equities and fixed-income can exhibit positive correlation for extended periods, most notably from 1984 through the late 1990s. Specific periods of market stress highlight significant shifts in this relationship; during the

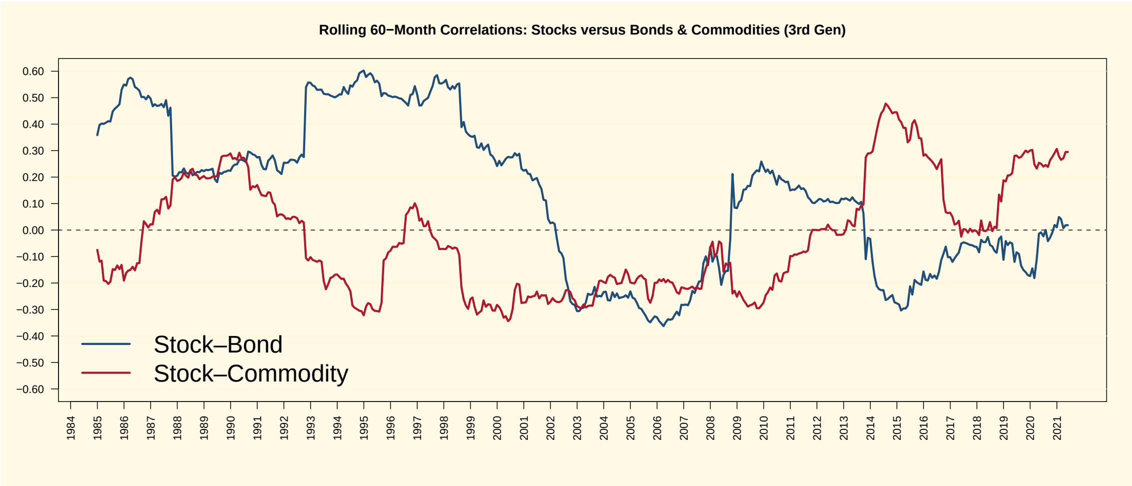

Figure 3: 60-month rolling correlation between stocks and the other two asset classes.

2008 financial crisis, the stock-bond correlation spikes sharply, moving from negative values (almost -0.30) into positive (+0.25). This indicates a temporary breakdown in diversification, as both asset classes moved in the same direction. A similar pattern emerges at the end of the sample period; approaching the 2020 COVID-19 crisis, the correlation trends upward again, crossing zero to become positive. This suggests that during extreme systemic shocks and heightened uncertainty, the typically inverse relationship between stocks and bonds tends to decouple, aligning with the literature (Brixton et al., 2023).

Figure 4 illustrates a persistent decline in the risk-free rate over the sample period. Starting from levels above 15% in the early 1980s, the rate trends downward, hitting the 0% level, following the 2008 financial crisis (quantitative easing) and again during the 2020 pandemic. The histogram confirms this concentration of values near zero, while the slowly decaying autocorrelation function indicates that the series is highly persistent and non-stationary.

The dividend-to-price ratio (Figure 5) exhibits significant long-term fluctuations

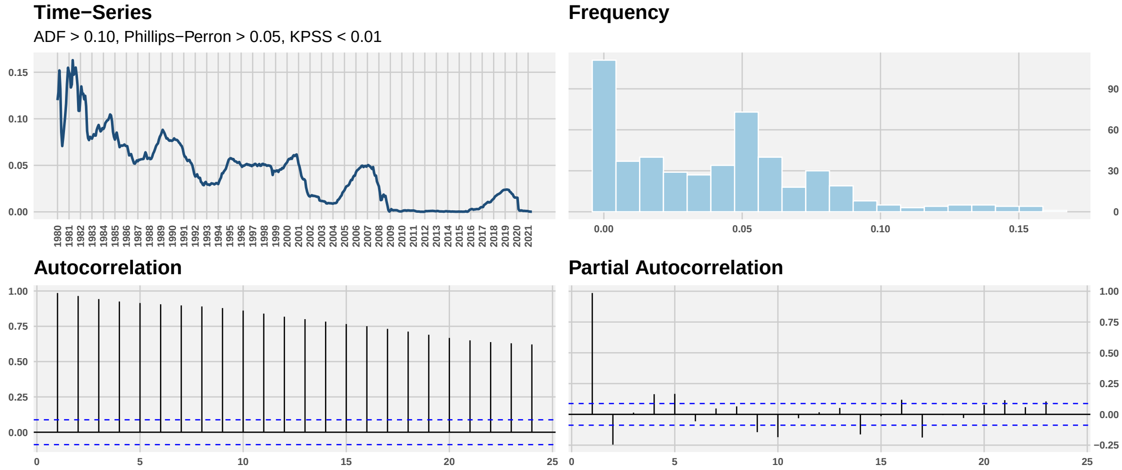

Figure 4: (Annualized) Treasury Bill Rate (TBL).

rather than a single linear trend. The ratio reached its lowest values during the late 1990s, reflecting the high market valuations of the dot-com bubble, and spiked during the 2008 financial crisis when stock prices collapsed. The series shows high persistence similar to interest rates, and the distribution reflects different regimes of market valuation over the last 40 years.

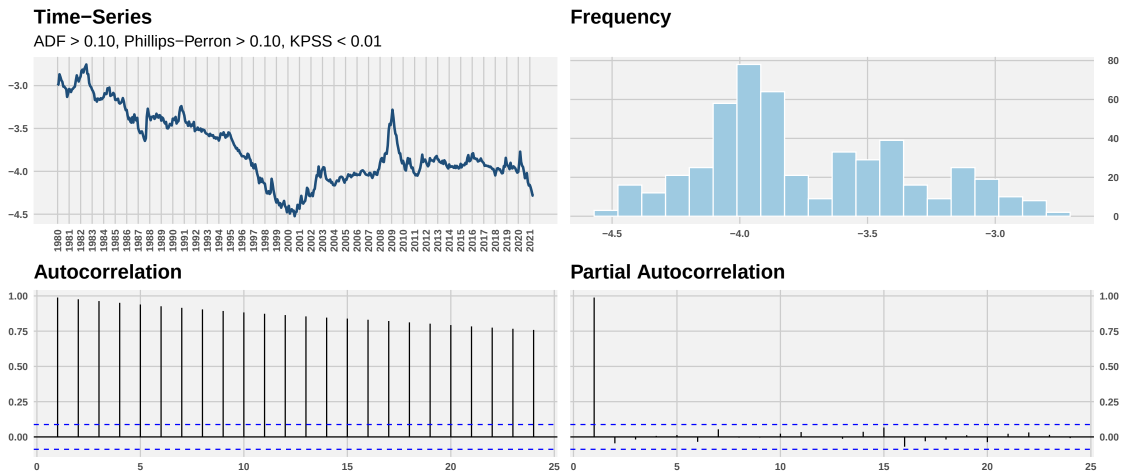

Figure 5: Dividend-to-price ratio (DP).

The term spread (Figure 6) displays a cyclical pattern that often corresponds with the business cycle. The series reverts to the mean more frequently than the

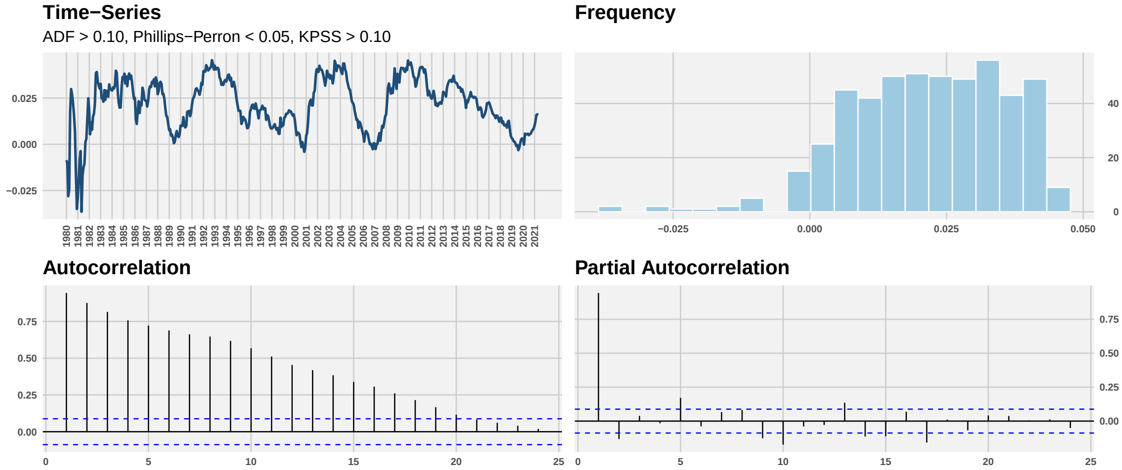

Figure 6: Term spread (TMS).

previous variables and includes notable periods where the spread drops below zero (yield curve inversion), typically preceding economic downturns.

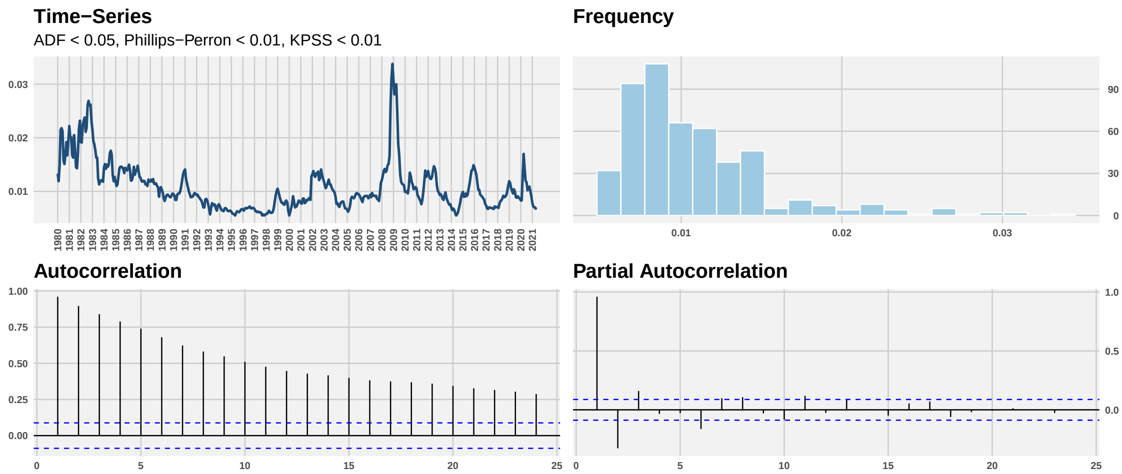

Figure 7: Default yield spread (DFY).

Finally, the default yield spread (Figure 7) serves as a proxy for credit risk, remaining relatively stable and low during economic expansions, but spiking sharply during stress events. The largest spike occurs during the 2008 financial crisis, with smaller spikes visible in the early 2000s and during the 2020 COVID-19 shock. The frequency distribution is highly right-skewed, confirming that while extreme credit stress is rare, it produces significant outliers.

While the statistical tests show that these variables are highly persistent (possible presence of unit roots), we do not focus on these statistical properties. Instead, our focus is on economic significance; the goal is to use this information to improve out-of-sample predictions in real time. If we can estimate expected returns even slightly better than simply using the historical average (unconditional mean), an investor can use that edge to build a better portfolio.

As a first exploratory step, Table 6 reports OLS estimates for the full sample to provide initial information into the direction of the relationships between the predictors and asset excess returns. Overall, the estimated signs are consistent with theoretical expectations and findings in the predictability literature. The Treasury Bill Rate (TBL) exhibits a negative association with future stock returns, while showing a positive relationship with bond and commodity returns. The dividend-to-price ratio is positively related to stock and commodity returns, but displays no clear relationship with bond returns. The term spread enters negatively for stocks and positively for both bonds and commodities. Finally, the default yield spread is negatively associated with stock returns and shows a weak or mixed (zero to positive) relationship with bond and commodity returns. Many of those slope coefficients are not statistically significant, underscoring the noisy nature of the estimates. Consistent with this, the adjusted R-Squared for bonds is close to zero in the full sample. Notably, restricting the sample to the 1982–2021 period yields an adjusted R-Squared of around 2% for bonds, reflecting the high volatility (around 12.5%) of bond returns during the early 1980–1982 period.

Table 6: Ordinary Least Squares (OLS) results for Stocks, Bonds, and Commodities over the full sample (1980-2021).

	Stocks		**Bonds**		**Commodities**	
	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
$Constant$	0.0072	3.69	0.0027	4.09	0.0046	3.57
	(0.0019)		(0.0007)		(0.0013)	
TBL_{t-1}	−0.0115	−3.36	0.0005	0.41	0.0084	3.68
	(0.0034)		(0.0012)		(0.0023)	
DP_{t-1}	0.0118	3.32	0.0003	0.23	−0.0073	−3.08
	(0.0036)		(0.0012)		(0.0024)	
TMS_{t-1}	−0.0040	−1.59	0.0014	1.59	0.0014	0.83
	(0.0025)		(0.0009)		(0.0017)	
DFY_{t-1}	−0.0031	−1.27	0.0007	0.77	0.0013	0.81
	(0.0025)		(0.0008)		(0.0016)	
$R^2_{adj.}(\%)$	1.84		0.00		2.67	

4 Methodology

Return predictability literature presents two contrasting perspectives that are highly relevant for portfolio choice. The skeptical view suggests that widely studied predictors (dividend–to-price ratio, dividend yield, earnings-to–price ratio, book-to-market ratio etc.) exhibit poor stability and limited out-of-sample predictive power, while the optimistic view emphasizes that carefully modeling conditional expected returns, by using theory driven constraints, can provide economically meaningful benefits to investors.

Goyal and Welch (2003) show that dividend ratios provide little reliable information about future stock market returns, with most of their predictive ability historically driven by a few isolated episodes, while dividend yield tend to forecast their own future values rather than expected returns. Expanding their analysis to a broader set of predictors, Goyal and Welch (2008) find that most forecasting models have performed poorly over the past several decades, offering little practical guidance for market timing or tactical asset allocation. Overall, their findings emphasize that in-sample predictability is often fragile and concentrated in exceptional periods (e.g., the 1973–1975 Oil Shock), and that unrestricted predictive models rarely outperform the historical mean in real-time applications.

In contrast, Campbell and Thompson (2008) argue that the poor performance of unrestricted predictive regressions does not imply a complete absence of predictability, but rather reflects the high estimation error in these models. They show that imposing theoretically motivated constraints such as enforcing non-negative expected returns improves out-of-sample performance. Even small gains in predictive accuracy translate into economically meaningful benefits, including higher SRs and improved investor utility, because conditional expected returns feed directly into the mean–variance optimization framework. These findings highlight that carefully modeling and estimating time-varying expected returns, guided by economic theory, can enhance portfolio choice and risk-adjusted performance.

Taken together, these studies suggest that while standard, unrestricted predictive regressions are often statistically fragile and unstable, theory disciplined models, could offer a more robust and economically meaningful approach to estimating conditional expected returns. Accordingly, although much of the literature emphasizes statistical metrics such as R^2, portfolio allocation should focus on the economic value of predictability. Even modest changes in conditional expected returns can materially influence optimal portfolio weights and enhance risk-adjusted performance.

4.1 Time Weighted Least Squares (TWLS)

We employ the time dependent weighted least squares (TWLS) (Hull et al. (2017), Wang et al. (2021)) model to address potential instability and structural breaks in the predictive relationship between predictors and asset returns. By assigning exponentially increasing weights to more recent observations, the model adapts more quickly to changes in the underlying data generating process, ensuring that structural shifts are incorporated faster than ordinary least squares (OLS). The linear model:

$$R_{t+1} = \boldsymbol{\beta' X_t} + u_{t+1}$$

where $\boldsymbol{X_t}$ is the $\boldsymbol{(K+1) \times 1}$ vector of predictors, and $\boldsymbol{\beta'}$ is the $\boldsymbol{1 \times (K+1)}$ vector with the model's slope estimates, while the R_{t+1} is the excess return, and u_{t+1} the error term, becomes:

$$\hat{\beta}_{TWLS} = \arg\min_{\beta} \sum_{t=1}^{T-1} \phi_t \left(R_{t+1} - \beta' X_t\right)^2$$

$$\phi_t = \lambda^{T-t}, \quad 0 < \lambda < 1$$

$$\text{Note that when } \lambda = 1 \quad \hat{\beta}_{TWLS} \text{ becomes identical with } \hat{\beta}_{OLS}$$

4.2 Exponentially Weighted Moving Covariance (EWMC)

To ensure consistency with the time-weighted estimation in the predictive regressions, we employ an exponentially time-weighted covariance matrix with a time-varying mean. The exponentially weighting scheme:

$$\psi_t = \frac{(1-\lambda)\,\lambda^{T-t}}{1-\lambda^T}, \qquad t = 1, \dots, T, \quad 0 < \lambda < 1$$

places greater emphasis on recent observations, so the estimated mean

$$\hat{\boldsymbol{\mu}}_{\boldsymbol{t}} = \sum_{t=1}^{T} \psi_t \boldsymbol{R}_{\boldsymbol{t}},$$

and covariance

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t}} = \sum_{t=1}^{T} \psi_t \, (\boldsymbol{R}_{\boldsymbol{t}} - \hat{\boldsymbol{\mu}}_{\boldsymbol{t}}) \, (\boldsymbol{R}_{\boldsymbol{t}} - \hat{\boldsymbol{\mu}}_{\boldsymbol{t}})^\top .$$

adapt more rapidly to potential structural breaks and short-term volatility spikes. This is consistent with the Time-Weighted Least Squares (TWLS) used in the previous step (which also upweights recent observations), ensuring that both conditional expected excess return estimates and risk (covariance matrix) estimates respond similarly to evolving market conditions. As a result, the inputs fed into a mean–variance optimization problem are more responsive to recent information and sensitive to current risk dynamics, producing portfolio allocations that better reflect the recent environment rather than relying on historical averages.

Figure 8 reports the time required to reach half-life (weight = 0.50) for different values of the smoothing parameter λ. Lower values of λ assign less weight to past observations, making the model more responsive to recent data. As $\lambda \to 1$, the weighting scheme converges to ordinary least squares (OLS) for the mean model, and the EWMC to the standard covariance matrix. In our baseline specification, we set $\lambda = 0.98$, which corresponds to a half-life of approximately 34 months. This choice implies that observations from the past three years receive relatively higher weights, while earlier periods have a weight below 0.50 and a more limited influence on the estimates. In the sensitivity analysis section, we examine the robustness of our results to alternative values of λ.

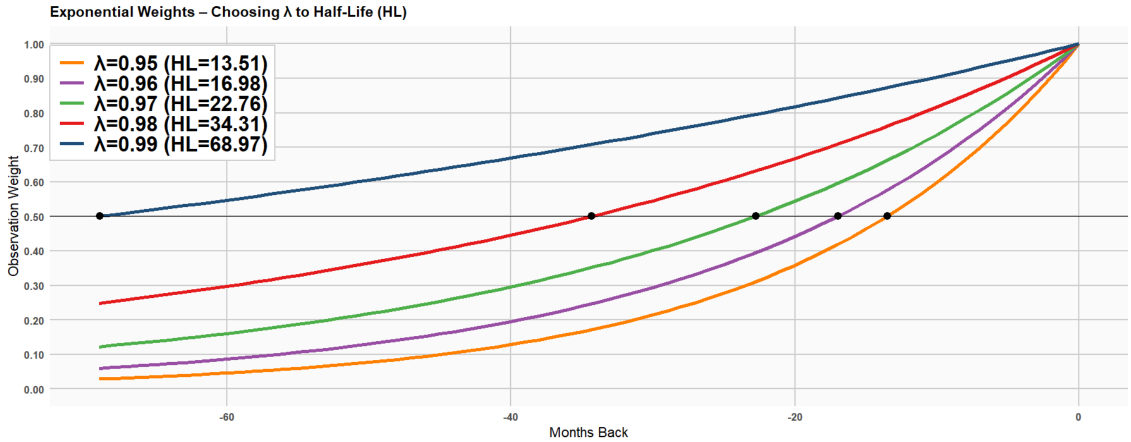

Figure 8: Months to half-life under different λ.

5 Results

As discussed in the data section, we use the Treasury bill rate (TBL), dividend-to-price ratio (DP), term spread (TMS), and default yield spread (DFY) to estimate the conditional expected excess returns for the three asset classes. The estimation is conducted using the time-weighted least squares approach explained earlier, with the smoothing parameter set to $\lambda = 0.98$. In addition, we impose sign constraints on the regression slope coefficients following economic intuition. The imposed constraints differ across asset classes and predictors, as summarized in Table 7.

Table 7: Slope constraints on predictive regressions.

Predictor	Stocks	Bonds	Commodities
Treasury Bill Rate (TBL)	Negative	Positive	Positive
Dividend-to-Price Ratio (DP)	Positive	No constraint	No constraint
Term Spread (TMS)	Negative	Positive	Positive
Default Yield Spread (DFY)	Negative	Positive	Positive

Estimation is performed using an expanding window that begins with an initial sample of 120 months (excess returns from February 1980 to January 1990, and predictors from January 1980 to December 1989, with the first realized return in February 1990). At each point in time, the dataset is updated to include newly available observations, and only information available up to time t is used. The estimated coefficients are then multiplied with the predictors observed at time t to obtain the expected excess return for period $t + 1$.

The slope coefficients for equities (Figure 9) offer support for the economic constraints imposed on the model. The negative constraints on the Treasury bill rate (TBL) and the term spread (TMS) appear to align naturally with the historical record, validating the discount rate channel; as rates or spreads rise, stock returns tend to fall. Although dividend-to-price ratio (DP) is constrained to be positive, the data reveals that this signal has largely faded. It was a powerful predictor in the early 1990s but essentially fell after 1998, suggesting that traditional valuation metrics lost their predictive power on pricing during the modern tech-driven era. Finally, the default yield spread (DFY) widens during periods of financial stress. Under the negative constraint, it remains low during normal market conditions, but spikes during crashes (e.g., 2001, 2008).

The bond market slope estimates (Figure 10) show time variation, although the pattern differs from that observed for equities. The coefficient on TBL is constrained to be positive, reflecting the expectation that higher short-term yields are associated with higher future bond returns. While this relationship holds throughout the sample, the coefficient displays instability around 1996 before stabilizing afterwards. The term spread coefficient, which is also restricted to be positive, suggests a shift in the mid-2000s, with the model assigning greater weight to the term premium than

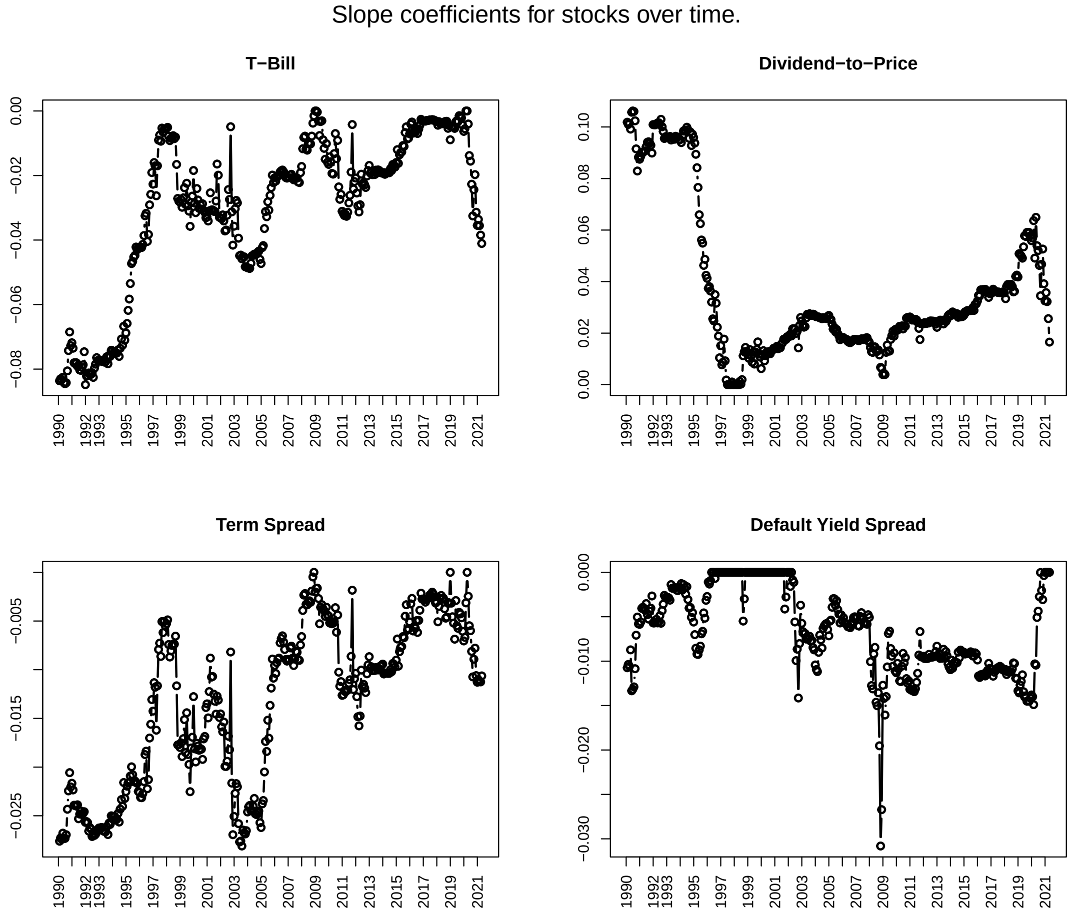

Figure 9: Model coefficients for stocks over time (1990 - 2021).

during the 1990s. In contrast, the dividend-to-price ratio is left unconstrained. Its coefficient gradually converges toward zero over the sample, mirroring the pattern observed in the equity market. This suggests that the decline in the predictive content of dividend yield during the late 1990s was not limited to equities, but reflected a broader change across asset classes.

Commodities, shown in Figure 11, display sensitivities to the macroeconomic variables that differ from those observed for equities and bonds. The Treasury bill (TBL) coefficient, which is constrained to be positive, remains relatively strong throughout the sample. This is consistent with the view that commodities provide

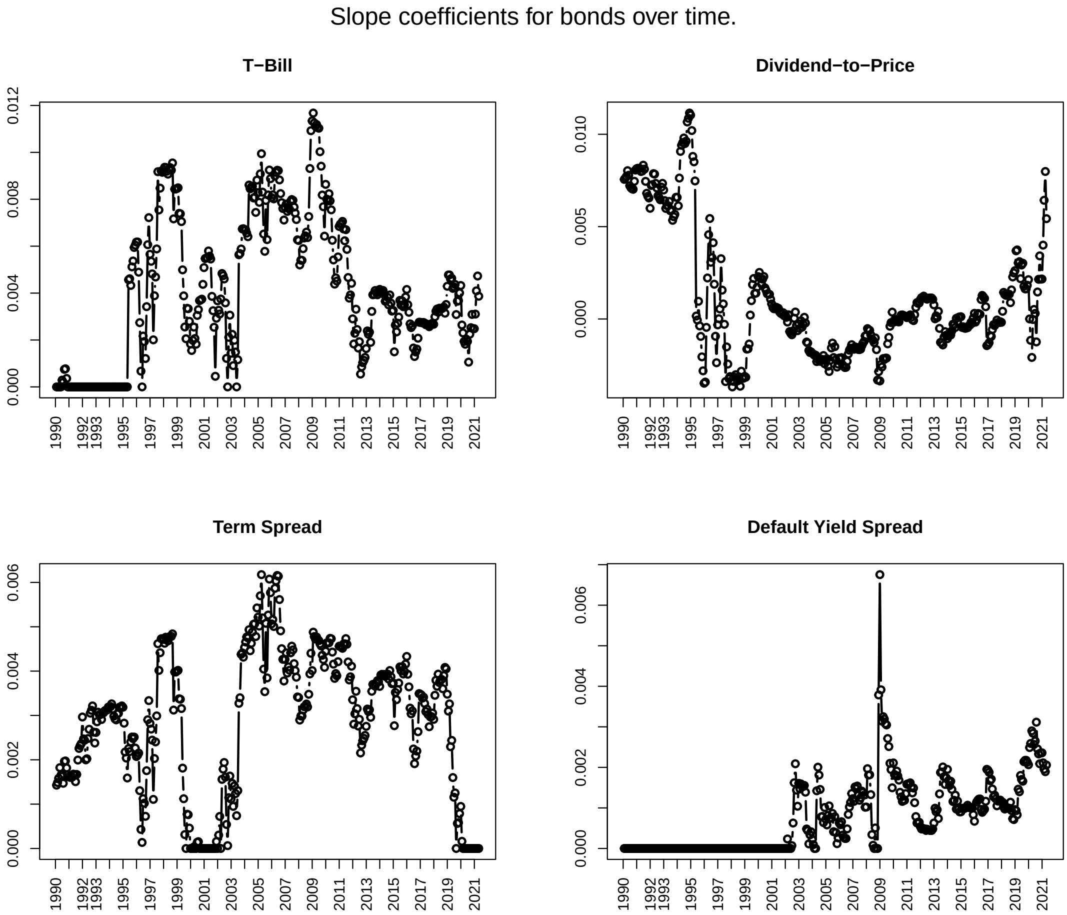

Figure 10: Model coefficients for bonds over time (1990 - 2021).

some protection against inflation, given the close relationship between short-term interest rates and inflation expectations. The coefficient on the term spread (TMS), also restricted to be positive, increases after 2000, indicating a stronger association between commodity returns and signals from the yield curve during the later part of the sample. The default yield spread (DFY), likewise constrained to be positive, reflects a risk-premium component that becomes more pronounced during periods of financial stress. During the 2008 liquidity crisis, the coefficient increases, suggesting that higher compensation was required to hold commodities under tightening credit conditions. In this respect, commodities behave more similarly to bonds than to equities, which experienced declines during the same period.

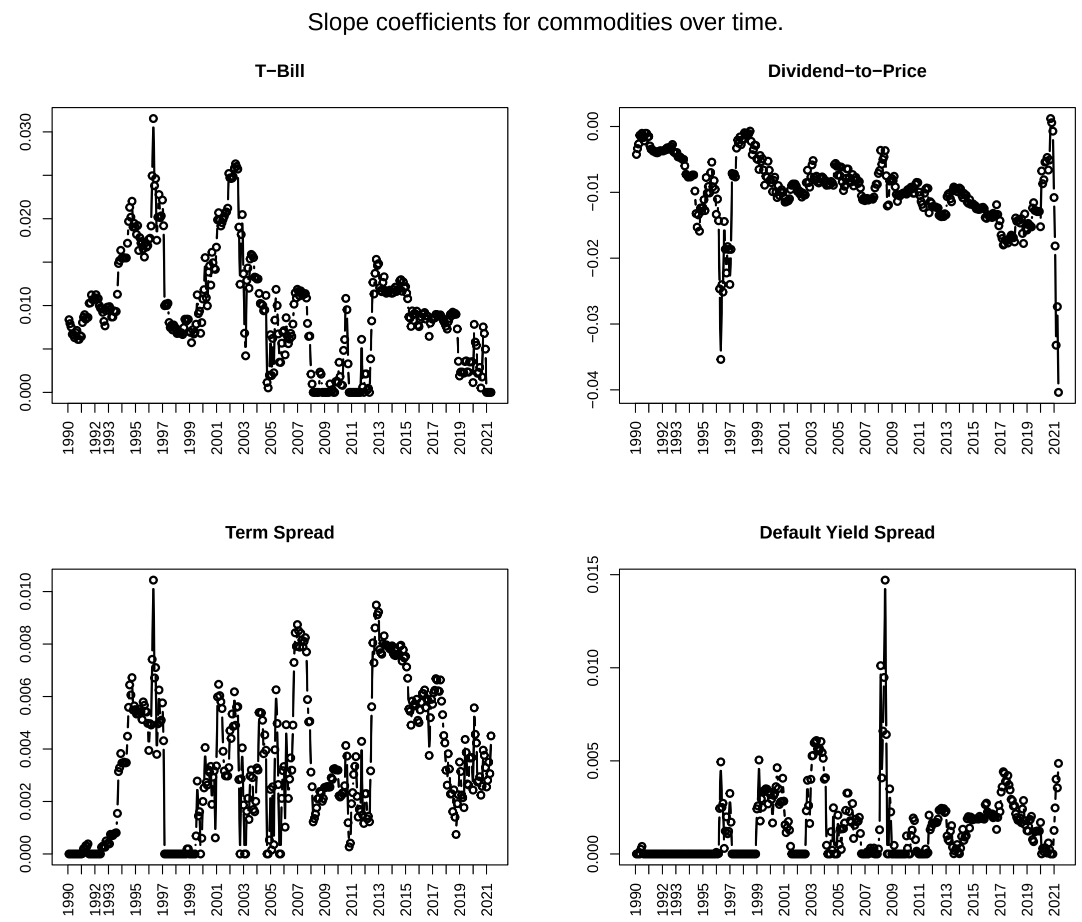

Figure 11: Model coefficients for commodities over time (1990 - 2021).

When viewed collectively, these plots demonstrate that while economic theory can correctly predict the direction of a relationship (hence, the validity of the sign constraints), it cannot predict the intensity, which changes over time. The use of standardized data allows us to see that the magnitude of these risks is not constant; for instance, the default yield spread (DFY) is largely inactive and becomes relevant during periods of systemic stress. A key takeaway is the breakdown of the dividend-to-price signal; whether constrained (Stocks) or unconstrained (Bonds/Commodities), the market moved away from this metric as a pricing factor in the late 1990s. Ultimately, these results provide evidence against the use of static models, as a constant slope would smooth over this variability and fail to capture key regime shifts.

5.1 Expected Returns and Risk

Figure 12 shows the conditional expected excess returns, which represent the model's estimation for what each asset will earn, while Table 8 demonstrates the properties of our estimates. Instead of relying on a single long-term average, these estimates evolve with the economic environment, with the conditional expected excess returns of equities, fixed income, and commodities generally remaining positive and persistent, while occasionally shifting across market cycles. The extreme observations from the 2007–2009 recession were removed to keep these outliers out for visualization purposes.

Table 8: Properties of the out-of-sample conditional expected excess returns.

Property	**Stocks**	**Bonds**	**Commodities**
Unconditional Mean (%)	9.76	1.83	3.87
Standard Dev. (%)	4.86	0.85	2.55
Positive Returns (%)	85.37	74.47	60.37
AR(1) Coef.	0.89	0.81	0.92

Description: Unconditional Mean (%) reports the unconditional mean of the conditional expected excess returns, and Standard Dev. (%) reports the corresponding annualized volatility. Positive Returns (%) indicates the proportion of estimates generating positive excess returns, and AR(1) Coef. refers to the coefficient of a first-order autoregressive model.

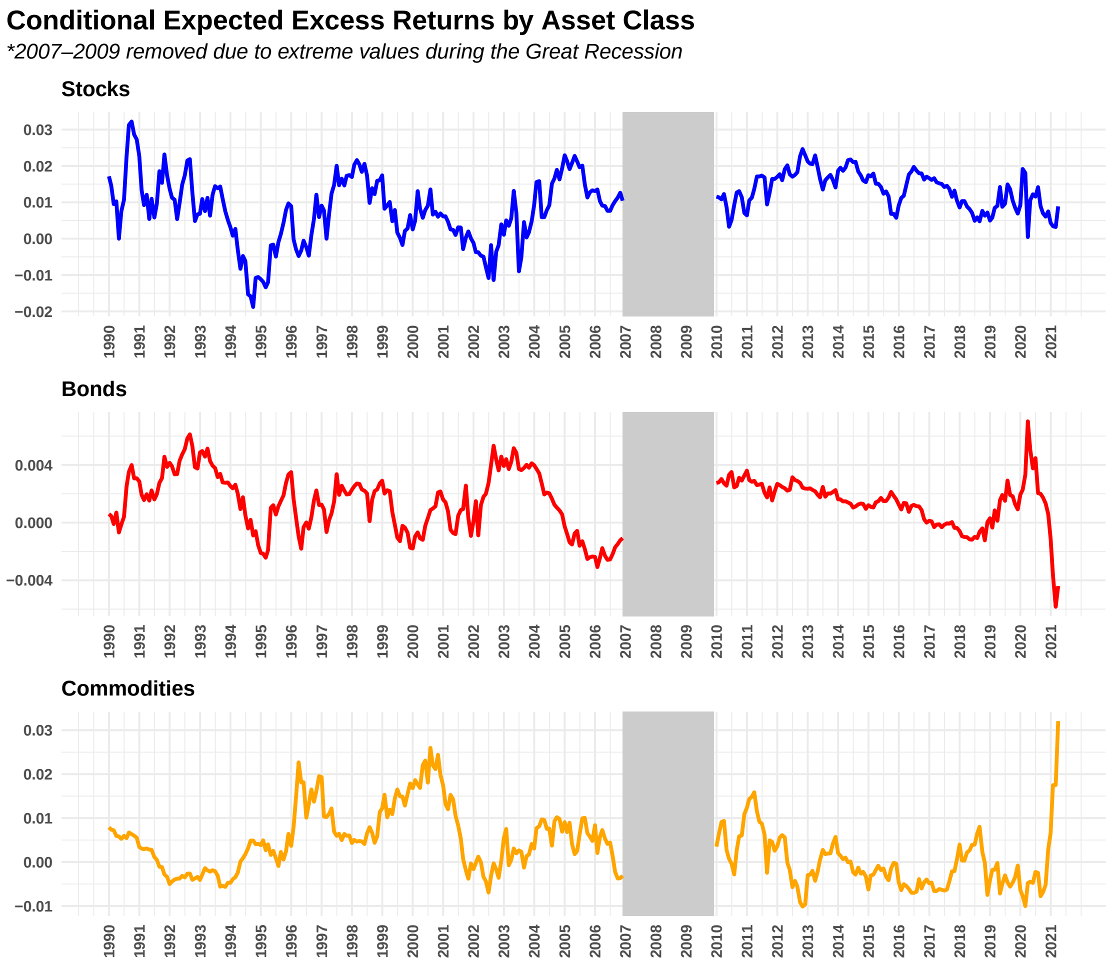

Figure 12: Out-of-Sample estimates for conditional expected excess returns ($\lambda = 0.98$).

Figure 13 shows the exponentially weighted variance, which serves as the risk input for the portfolio model. Volatility is calculated using an exponentially weighted mean rather than a historical average, consistent with the EWMC framework described in Section 4.2. This ensures that risk is measured as the deviation from the model's current prediction, not a long-term trend. By using exponential weighting, the calculation prioritizes recent data, allowing the risk estimate to adjust significantly faster when market volatility increases.

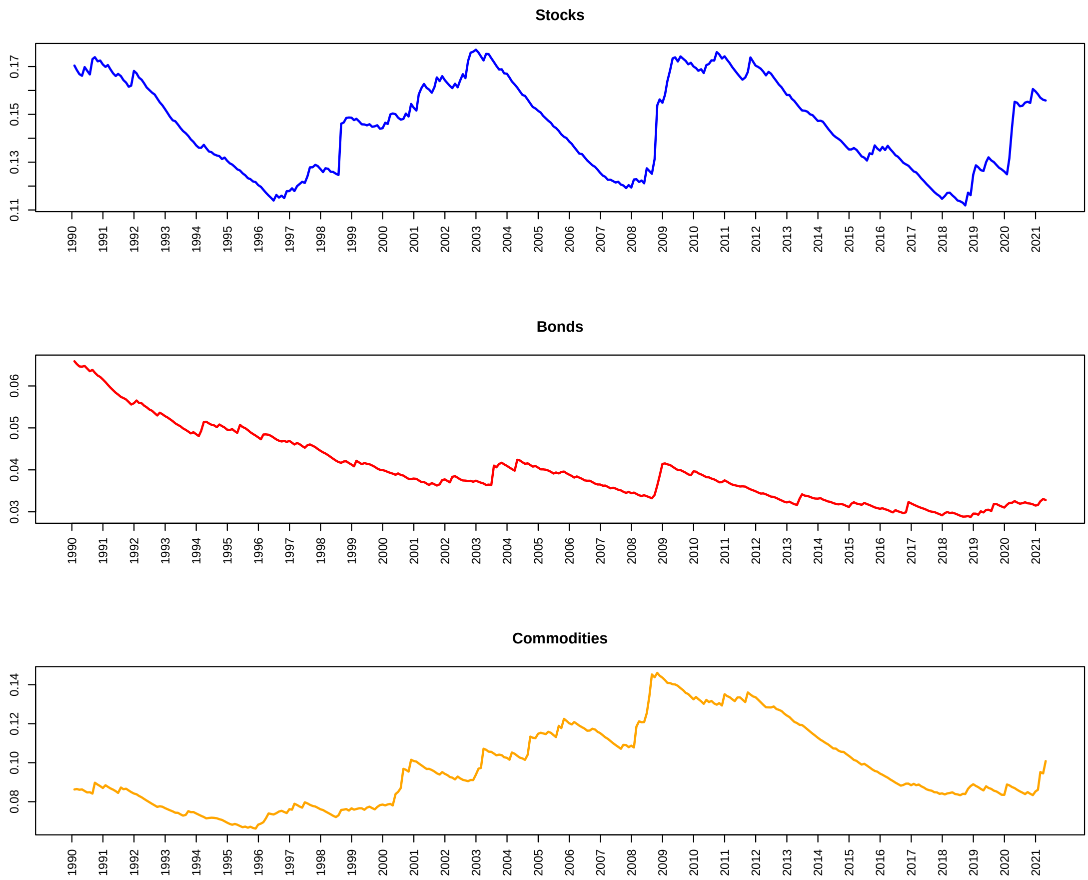

Figure 13: Exponentially weighted (annualized) volatility ($\lambda = 0.98$).

5.2 Mean-Variance Optimization

We consider an investor with mean–variance preferences, whose utility function is given by:

$$U = \mu - \frac{\gamma}{2}\sigma^2$$

where μ denotes the portfolio's expected excess return, σ^2 the portfolio's return variance, and γ the coefficient of risk aversion. The investor chooses portfolio weights to maximize his expected utility. In the baseline specification, we set the coefficient of risk aversion to $\gamma = 3$, corresponding to a reasonable level of risk-aversion. In the sensitivity analysis section, we examine the robustness of the results to alternative values of γ.

In our case, the mean–variance framework is extended to a multi-asset setting. Let $\boldsymbol{\mu_t}$ denote the vector of conditional expected excess returns and $\boldsymbol{\Sigma_t}$ the corresponding covariance matrix. The investor chooses portfolio weights $\mathbf{w_t}$ to maximize expected utility, given by:

$$U(\mathbf{w}_t) = \mathbf{w}_t^\top \boldsymbol{\mu}_t - \frac{\gamma}{2}\,\mathbf{w}_t^\top \boldsymbol{\Sigma}_t \mathbf{w}_t$$

We consider four portfolio specifications. The first two models assume no return predictability: (i) a portfolio consisting of stocks and bonds, and (ii) a portfolio consisting of stocks, bonds, and commodities (henceforth NPSB for No Predictability Stock-Bond and NPSBC for No Predictability Stock-Bond-Commodity), and their mean-variance inputs are also estimated with respect on time variation (time weighted average for mean and the same EWMC matrix for variance). The remaining two models incorporate the proposed predictive strategy: (iii) a stock–bond portfolio (PSB: Predictability Stock-Bond), and (iv) a stock–bond–commodity portfolio (PSBC: Predictability Stock-Bond-Commodity). Additionally, portfolio choice is subject to several constraints: leverage is allowed, with the sum of portfolio weights equals to the maximum of 1.5. Short selling is not allowed and in the three asset case, each asset weight is constrained to lie between $\underline{\mathrm{w}}_{\mathrm{i}} = 5\%$ and $\overline{\mathrm{w}}_{\mathrm{i}} = 90\%$, while in the two asset case, the minimum weight is set to $\underline{\mathrm{w}}_{\mathrm{i}} = 10\%$, with the same upper bound of $\overline{\mathrm{w}}_{\mathrm{i}} = 90\%$. Hence, the problem becomes:

$$\max_{\mathbf{w}_t} \quad \mathbf{w}_t^\top \boldsymbol{\mu}_t - \frac{\gamma}{2} \mathbf{w}_t^\top \boldsymbol{\Sigma}_t \mathbf{w}_t$$

$$\text{s.t.} \quad \mathbf{1}^\top \mathbf{w}_t \leq 1.5, \quad \underline{\mathbf{w}}_\mathbf{i} \leq \mathbf{w}_{\mathbf{t},\mathbf{i}} \leq \overline{\mathbf{w}}_\mathbf{i}$$

In Table 9, we summarize the out-of-sample performance. The results confirm that expanding the asset universe to include commodities provides meaningful diversification benefits, regardless of whether return predictability is incorporated. Notably, the NPSBC strategy already achieves a SR of 0.94 and a CAGR of 13.33%, outperforming the PSB strategy, suggesting that commodity diversification alone, even without predictability, adds value. These benefits become even more pronounced when predictability is combined with commodity allocation; the PSBC strategy delivers the highest terminal wealth while maintaining a SR of 0.95. This highlights that while commodities are the primary driver of this outperformance, the predictive framework amplifies the gains in the three-asset case.

Table 9: Out-of-sample portfolio performance. Strategies with no return predictability (NP) versus strategies with return predictability (P).

STRATEGY	CAGR	AV	SD	SR	SORT	MDD	W
No Predictability (NP)							
Stock–Bond (SB)	11.47	8.88	10.57	0.84	1.13	19.33	30.03
Stock–Bond–Commodity (SBC)	13.33	10.64	11.37	0.94	1.43	18.43	50.41
Predictability (P)							
Stock–Bond (SB)	11.75	9.19	11.13	0.83	1.16	19.13	32.47
Stock–Bond–Commodity (SBC)	14.27	11.56	12.12	0.95	1.33	21.85	65.38

Description: **(1)** CAGR: Compound Annual Growth Rate (%), **(2)** AV: Annualized Excess Return (%), **(3)** SD: Annualized Volatility (%), **(4)** SR: Sharpe Ratio, **(5)** SORT: Sortino Ratio, **(6)** MDD: Maximum Drawdown (%), **(7)** W: Terminal Wealth.

In addition to these performance measures, we evaluate each portfolio using the

Certainty Equivalent Return (henceforth CER). The CER represents the guaranteed return an investor would accept instead of facing the risky portfolio, given their risk aversion, and provides a single measure that accounts for both return and risk. Table 10 reports CERs for different risk aversion coefficients (γ) across the four portfolio specifications considered: No Predictability (NP) and Predictability (P) based allocations, with two asset (SB: stocks–bonds) and three asset (SBC: stocks–bonds–commodities) portfolios. As expected, CER declines with increasing risk aversion, reflecting the greater penalization of risk. Consistently, the three-asset strategies deliver higher CERs than their corresponding stock-bond benchmarks, with the largest gains observed in the predictability framework (PSBC strategy). These results complement the earlier measures (Table 9), highlighting the benefits of predictive signals and broader diversification for improving risk-adjusted portfolio performance.

Table 10: Certainty Equivalent Returns (CER) by risk aversion coefficient. No Predictability vs Predictability Portfolios.

	NPSB	**NPSBC**	**PSB**	**PSBC**
$\gamma = 1$	10.91	12.57	11.16	13.41
$\gamma = 2$	10.35	11.93	10.54	12.68
$\gamma = 3$	9.79	11.28	9.93	11.95
$\gamma = 4$	9.24	10.63	9.31	11.21
$\gamma = 5$	8.68	9.99	8.70	10.48

Description: CE Returns (in %) for different risk aversion coefficients. NP: No Predictability Strategy Portfolio, P: Predictability Strategy Portfolio, SB: Stock-Bond, SBC: Stock-Bond-Commodity.

We observe that the PSBC strategy achieves the highest CER for every level of γ. The sensitivity analysis (Tables 11 & 12), which reports the SR and CAGR for varying γ and λ values, confirms the three-asset case dominance. For higher values of λ, PSBC outperforms NPSBC. In contrast, for lower values of λ, NPSBC tends to perform better.

Table 11: Out-of-sample portfolio performance (Sharpe ratios).

Strategy	λ	γ = 1	γ = 2	γ = 3	γ = 4	γ = 5
NPSB	0.95	0.85	0.88	0.87	0.88	0.89
	0.96	0.84	0.84	0.88	0.89	0.89
	0.97	0.83	0.86	0.86	0.88	0.88
	0.98	0.81	0.82	0.84	0.84	0.85
	0.99	0.78	0.77	0.78	0.80	0.81
NPSBC	0.95	0.94	0.97	0.99	0.99	0.99
	0.96	0.93	0.96	0.98	0.99	0.98
	0.97	0.92	0.94	0.97	0.98	0.98
	0.98	0.91	0.92	0.94	0.94	0.94
	0.99	0.81	0.84	0.84	0.84	0.85
PSB	0.95	0.78	0.82	0.83	0.83	0.84
	0.96	0.82	0.83	0.82	0.84	0.84
	0.97	0.81	0.82	0.84	0.84	0.85
	0.98	0.80	0.82	0.83	0.85	0.85
	0.99	0.75	0.76	0.77	0.79	0.81
PSBC	0.95	0.91	0.95	0.95	0.94	0.95
	0.96	0.96	0.96	0.95	0.94	0.94
	0.97	0.97	0.99	0.99	0.98	0.98
	0.98	0.95	0.95	0.95	0.97	0.96
	0.99	0.88	0.87	0.88	0.90	0.90

Notes: Each cell reports **Sharpe Ratio**. λ denotes the exponential weighting factor and γ the coefficient of risk aversion.

This pattern suggests that within the OLS based framework, relative to the historical average benchmark, incorporating predictability through macroeconomic variables can be beneficial for the mean–variance investor. Conversely, when the model primarily accounts for time variation in the inputs, the additional information from macroeconomic variables appears less critical, as much of this information may already be reflected in asset prices. In such cases, placing greater weight on recent observations ($\lambda = 0.95$ implies that only observations from approximately the last 12 months receive weights above 0.50) can effectively capture the current market environment. However, lower values of λ introduce a practical drawback. Because the mean–variance inputs adjust rapidly to new information, portfolio weights can change substantially from month over month, leading to frequent rebalancing. This may result in significant transaction costs, which could ultimately offset the performance gains implied by the model.

Table 12: Out-of-Sample Portfolio Performance (CAGR %).

Strategy	$\boldsymbol{\lambda}$	$\boldsymbol{\gamma = 1}$	$\boldsymbol{\gamma = 2}$	$\boldsymbol{\gamma = 3}$	$\boldsymbol{\gamma = 4}$	$\boldsymbol{\gamma = 5}$
NPSB	0.95	11.69	11.76	11.42	11.14	10.92
	0.96	11.78	11.55	11.61	11.43	11.01
	0.97	11.81	11.78	11.61	11.47	11.01
	0.98	11.77	11.61	11.47	11.18	10.76
	0.99	11.96	11.56	11.04	10.78	10.25
NPSBC	0.95	13.61	13.80	13.72	13.45	13.14
	0.96	13.57	13.71	13.76	13.53	13.22
	0.97	13.56	13.60	13.76	13.57	13.23
	0.98	13.54	13.38	13.33	13.06	12.73
	0.99	12.35	12.52	12.21	11.72	11.38
PSB	0.95	10.91	11.15	11.02	10.81	10.57
	0.96	11.67	11.48	11.16	11.01	10.85
	0.97	11.83	11.73	11.58	11.36	11.19
	0.98	11.81	11.90	11.75	11.69	11.38
	0.99	11.16	11.09	11.07	11.12	10.94
PSBC	0.95	13.50	13.86	13.63	13.35	13.15
	0.96	14.28	14.11	13.87	13.59	13.32
	0.97	14.58	14.69	14.41	14.13	13.78
	0.98	14.42	14.34	14.27	14.17	13.80
	0.99	13.35	13.13	13.11	13.19	13.05

Notes: Each cell reports **CAGR (%)**. λ denotes the exponential weighting factor and γ the coefficient of risk aversion.

5.3 Statistical Significance of the Sharpe Ratio (SR) and the Certainty Equivalent Return (CER)

To determine whether this observed outperformance of the PSBC strategy over the PSB strategy is statistically significant, we employ a formal hypothesis testing framework. While Jobson and Korkie (1981) developed the traditional test for the equality of two SRs, their methodology relies on the strict assumption that returns are independent and identically distributed (I.I.D.) with a normal distribution. However, financial data frequently have heavy tails (non-normality) and volatility clustering (serial correlation), which make standard normal tests invalid. To address these empirical properties, we adopt a robust asymptotic inference framework. Following the specific implementation in Jiang et al. (2018), we utilize the asymptotic delta method rather than a bootstrap procedure (Ledoit and Wolf, 2008). This approach relies on Heteroskedasticity and Autocorrelation (HAC) standard errors. Our testing procedure involves the following steps:

Sharpe Ratio (SR) Test: We test the null hypothesis $H_0 : SR_{PSBC} = SR_{PSB}$ against the one-sided alternative $H_1 : SR_{PSBC} > SR_{PSB}$. Since the SR is a non-linear function of the mean and variance, we utilize the delta method to approximate the variance of the SR difference. We estimate the covariance matrix using the Newey and West (1987) estimator (with a maximum lag of 4).

Certainty Equivalent Return (CER) Test: We extend this methodology to the CER; we test the null hypothesis $H_0 : CER_{PSBC} = CER_{PSB}$ against the one-sided alternative $H_1 : CER_{PSBC} > CER_{PSB}$. Although literature primarily focuses on the SR, the same statistical properties apply to utility-based metrics. We define the utility gain as the difference in CER between the two strategies for a mean-variance investor with a degree of risk-aversion $\gamma = 3$. Using again the delta method for the CER function, and applying the Newey and West (1987) correction, we test whether the utility provided by the three asset portfolio (PSBC) is statistically higher from the two asset benchmark (PSB).

For the SR, the one-sided test yields a p-value between 0.10 and 0.15. This indicates that while the point estimate of the risk-adjusted return is higher for the three asset strategy, the difference is not statistically significant at conventional levels (10% or 5%). This is likely due to the high variability associated with the estimator in finite samples. However, the results for the CER are stronger. For an investor with a risk aversion coefficient of $\gamma = 3$, the test yields a p-value around 5%, implying that the utility gain provided by adding commodities is statistically significant at the 10% level. This divergence suggests that while the reward-to-variability ratio (Sharpe) may not be greater in statistical terms, the structural improvement in the portfolio's utility profile (CER) is robust enough to be distinguished from noise.

5.4 First, Second, and Third Generation of Commodities

As discussed in the literature review section, the diversification benefits of commodities are primarily associated with third-generation commodity indices. After evaluating the strategy based on the Morningstar Long/Short Commodity Index (third-generation), we apply the same methodology and framework to a first-generation index (S&P GSCI) and a second-generation index (Morningstar Long/Flat). The out-of-sample results are reported in Table 13.

When incorporating the second-generation commodity index, the SR remains unchanged relative to the stock–bond benchmark, while terminal wealth increases modestly and maximum drawdown remains relatively low. In contrast, the inclusion of the first-generation index does not provide diversification benefits. Its terminal wealth is the same with the stock–bond benchmark, but accompanied by substantially higher drawdowns and a fall in the SR, driven by increased portfolio volatility.

The strongest improvement is observed when using the third-generation commodity index, which in this case, nearly all performance measures improve, including higher returns and superior risk-adjusted performance. Although statistical significance cannot be established for the SR at the 10% level, the economic relevance of these improvements is evident. From a mean–variance perspective, many investors with a reasonable degree of risk aversion would likely prefer the three asset class portfolio (PSBC), given its higher terminal wealth without changing the drawdown profile, as illustrated in Figures 14 and 15.

Table 13: Out-of-Sample Performance: Strategy with predictability of the stock-bond portfolio (PSB) versus the results with the addition of the three generations of commodity indices.

	Stock–Bond	**1st Gen**	**2nd Gen**	**3rd Gen**
CAGR (%)	11.75	11.97	12.96	14.27
Premium (%)	9.19	9.72	10.44	11.56
Volatility (%)	11.13	13.72	12.47	12.12
Sharpe	0.83	0.71	0.84	0.95
Sortino	1.16	0.99	1.13	1.33
Max Drawdown (%)	19.13	30.42	22.89	21.85
Terminal Wealth	32.47	34.54	45.52	65.38

Description: **CAGR**: Compound Annual Growth Rate (%), **Premium**: Annualized Excess Return (%), **Volatility**: Annualized Standard Deviation (%), **Sharpe**: Sharpe Ratio, **Sortino**: Sortino Ratio, **Max Drawdown**: Maximum Drawdown (%), **Terminal Wealth**: Final portfolio value of an $1 initial investment.

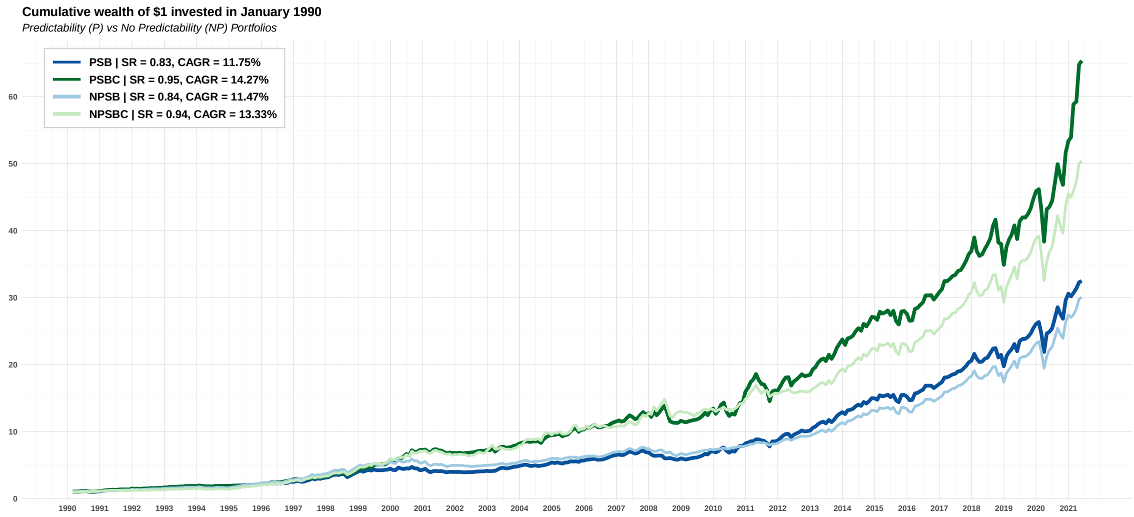

Figure 14: Wealth paths of different alternatives over time (1990 - 2021).

Figure 14 shows the wealth paths of the different alternatives, while Figure 15 shows the drawdowns. For both figures, the commodities we refer to are from the third-generation index (Morningstar Long/Short).

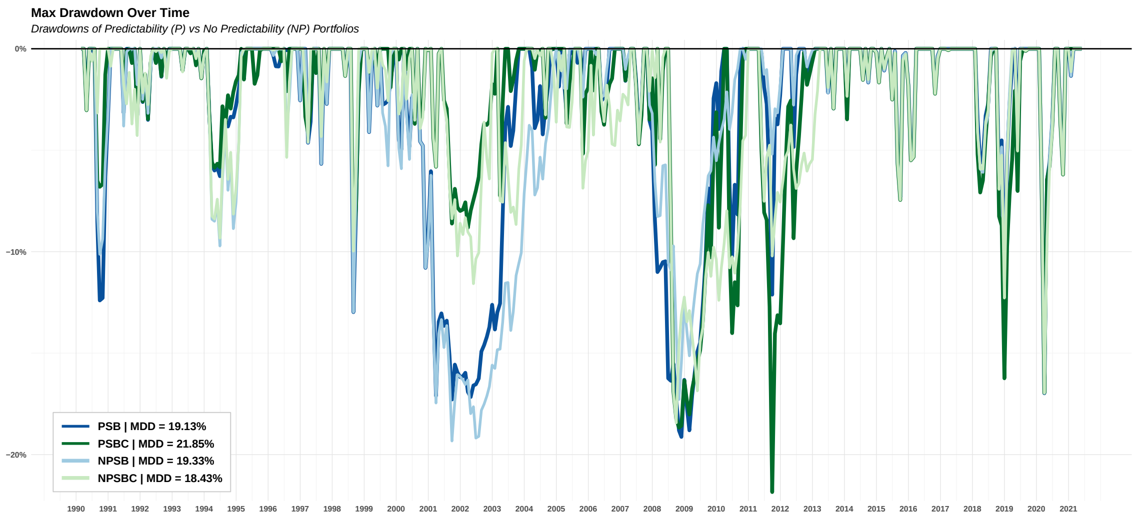

Figure 15: Maximum drawdown of different alternatives over time (1990 - 2021).

6 Conclusions

This study investigated whether commodities could improve the performance of traditional equity-bond portfolios. Rather than relying on static allocations (e.g., the 60/40) or unconditional historical averages, we developed a dynamic out-of-sample investment strategy that adapts to structural breaks in the economy. By employing Time-Weighted Least Squares (TWLS) and imposing theoretically motivated sign constraints on the slope coefficients, we utilized four macroeconomic predictors to estimate conditional expected returns and covariance matrices in real-time. We tested this framework across three generations of commodity indices to isolate the impact of index construction, and our findings suggest that while first-generation (long-only, production-weighted) indices do not add value, third-generation (long/short, factor-based) indices can provide diversification benefits, aligning with existing literature. The significance of these findings lies in the economic value delivered to the mean-variance investor. Our analysis confirms that the correlation between stocks and bonds is time-varying and subject to reversals, particularly during periods of economic stress, where the diversification benefits of a traditional equity-bond portfolio diminish. The results also provide evidence for return predictability. While the predictive model (PSBC) does not consistently outperform the no-predictability benchmark (NPSBC) in risk-adjusted terms (NPSBC performs better for lower values of λ), the imposition of sign constraints serves as a useful tool for preventing overfitting, allowing the predictability strategy (PSBC) to generate the highest terminal wealth. The model successfully activates the default yield spread (DFY) signal during periods of market stress and correctly identifies the fading of the dividend-to-price ratio in the late 1990s aligning with Goyal and Welch (2003). A static model would have continued to rely on this relationship, leading to poor performance. However, our adaptive TWLS model correctly identified that the signal had faded and naturally reduced its influence, demonstrating the robustness of the methodology against structural breaks. The primary source of outperformance in this framework appears to be commodity diversification rather than return predictability. The addition of third-generation commodity indices delivers the strongest results, consistent with Miffre (2012) and Fethke and Prokopczuk (2018).

Finally, this thesis takes a practical stance to the allocation problem. Rather than relying on full-sample data that investors could not have known in advance, we address the asset allocation problem as faced by hedge fund managers and investors in real time. By constructing portfolio weights that adapt to the business cycle, we show that the diversification value of commodities is not static. Third-generation indices, through their dynamic long/short construction, deliver robust and persistent out-of-sample diversification benefits.

7 Limitations and Discussion

While the findings of this study suggest the inclusion of third-generation commodity indices in a dynamic portfolio, certain limitations must be acknowledged regarding the implementation of the strategy. First, any active strategy that involves monthly rebalancing incurs transaction costs, which are not deducted in our calculations. However, in practice, this concern is manageable. The strategy allocates capital across only three broad asset classes, which can be accessed through liquid Exchange Traded Funds (ETFs) with relatively low expense ratios. Furthermore, the use of a smoothing parameter ($\lambda = 0.98$) in our predictive model ensures that the estimated conditional returns and consequently the portfolio weights evolve gradually over time. This stability prevents excessive turnover and dramatic month-over-month shifts in allocation, keeping trading friction within acceptable limits. A more significant limitation lies in the accessibility of the specific financial instruments required to replicate the third-generation commodity performance. Unlike first-generation indices, which are widely available to the public through standard ETFs, third-generation indices that employ dynamic Long/Short strategies are complex and often difficult for individual investors to locate in the retail market. This creates a barrier to entry for the ordinary investor, who may face challenges finding a product that mirrors an index like the one used in this study. However, this limitation is less restrictive for hedge fund managers and institutional investors. These participants have the infrastructure capability to create such positions, allowing them to fully capitalize on the strategic advantage identified in this thesis.

Future research should move beyond index level analysis and focus on identifying specific investable products. While indices provide a theoretical benchmark, mapping these returns to actual Exchange Traded Products (ETPs) or notes available in the market would allow for practical implementation. Second, while this thesis utilized a common set of four macroeconomic predictors across all asset classes, future models could benefit from tailoring predictors to each specific asset. The drivers of commodity returns often differ from those of financial assets; therefore, incorporating asset specific predictors could enhance the forecasting accuracy of the TWLS model. Finally, extending the dataset beyond May 2021 (which could not be done as the Morningstar Long/Short was terminated in 2021) would be of significant interest. The period from 2021 to 2025 represents an economic regime characterized by the boom in Gold and Silver prices, heightened inflation uncertainty, significant geopolitical tensions and wars. These are the conditions under which commodities are theorized to outperform. Testing the dynamic strategy during this post-pandemic era would provide a robust out-of-sample stress test, potentially confirming the strategy's value as a hedge against inflation and geopolitical risk in a modern context. In the Appendix, we introduce ETF proxies that allow us to extend the analysis to the 2021–2024 period, and in Section A1 we show that our results continue to hold in the post-pandemic period and are even stronger.

8 Appendix

While this thesis utilizes historical index data to evaluate the performance of different asset allocation strategies, investors could potentially implement those strategies in real-time using Exchange-Traded Funds (ETFs) that track the respective indices. For traditional assets, the *State Street SPDR S&P 500 ETF Trust (SPY in yahoo)* and *iShares Core U.S. Aggregate Bond ETF (AGG in yahoo)* provide direct exposure to the equity and bond benchmarks used in this study. First-generation commodity exposure can be achieved through the *iShares S&P GSCI Commodity-Indexed Trust (GSG in yahoo)*. While second and third generation indices demonstrate superior historical performance in this research, direct tracking products for the specific Morningstar indices used are currently limited as the indices were terminated in 2021. However, alternatives such as the *Direxion Auspice Broad Commodity Strategy ETF (COM in yahoo)* and the *Simplify Commodities Strategy No K-1 ETF (HARD in yahoo)* employ similar dynamic roll and long/short methodologies which could be used for second and third generation indices. We extend our dataset using these ETFs, to review how each strategy performs until December 2024.

A1 Extended sample period

This subsection presents a robustness check based on the extended sample period. In the main analysis, the sample spans from January 1980 to May 2021, reflecting the inception and termination of the Morningstar indices. To extend the analysis beyond this window, we construct a composite commodity index using the available proxies across subperiods. Specifically, the COM ETF (proxy for a second-generation commodity index) is used for June 2021 to March 2023, while the HARD ETF (proxy for a third-generation commodity index) is employed from April 2023 (inception date) to December 2024. We replicate our methodology using this extended sample, thereby providing an additional robustness check and allowing us to validate whether the results continue to hold in the post-COVID period. Notably, the inclusion of the 2021–2024 period strengthens the documented benefits of commodity allocation; this is driven by a decline in the SR of the traditional stock-bond portfolio, resulting from heightened stock-bond correlation and increased macroeconomic uncertainty during the extended period. The statistical tests confirm the significance of these improvements, as the p-values for the differences in SRs remain below 10% (across maximum 3–6 lags), while the differences in CER are also significant with p-values below 5% (across maximum 3–6 lags). These findings demonstrate that the diversification benefits of commodities not only persist but are amplified during periods of structural shifts (wars, geopolitical tensions, political uncertainty etc.) and market volatility, as explained in the literature review section. In Table A.1, we present the performance measures for the extended period (1980 - 2024), covering 34 years (1990-2024) in out-of-sample realized returns, and in Figure A.1 the wealth paths of the four proposed strategies. This sample extension let us obtain statistically significant improvements in both SR and CER, alongside uniformly superior performance across all reported measures, using the best alternative data we can find.

Table A.1: Out-of-sample portfolio performance. Strategies with no return predictability (NP) versus strategies with return predictability (P). *(Extended Sample)*

STRATEGY	**CAGR**	**AV**	**SD**	**SR**	**SORT**	**MDD**	**W**
No Predictability (NP)							
Stock–Bond (SB)	11.20	8.63	11.22	0.77	1.04	26.68	40.73
Stock–Bond–Commodity (SBC)	13.06	10.39	11.94	0.87	1.31	23.13	72.63
Predictability (P)							
Stock–Bond (SB)	11.07	8.46	10.81	0.78	1.10	19.13	39.05
Stock–Bond–Commodity (SBC)	13.89	11.12	11.97	0.93	1.32	21.85	93.70

Description: **(1)** CAGR: Compound Annual Growth Rate (%), **(2)** AV: Annualized Excess Return (%), **(3)** SD: Annualized Volatility (%), **(4)** SR: Sharpe Ratio, **(5)** SORT: Sortino Ratio, **(6)** MDD: Maximum Drawdown (%), **(7)** W: Terminal Wealth.

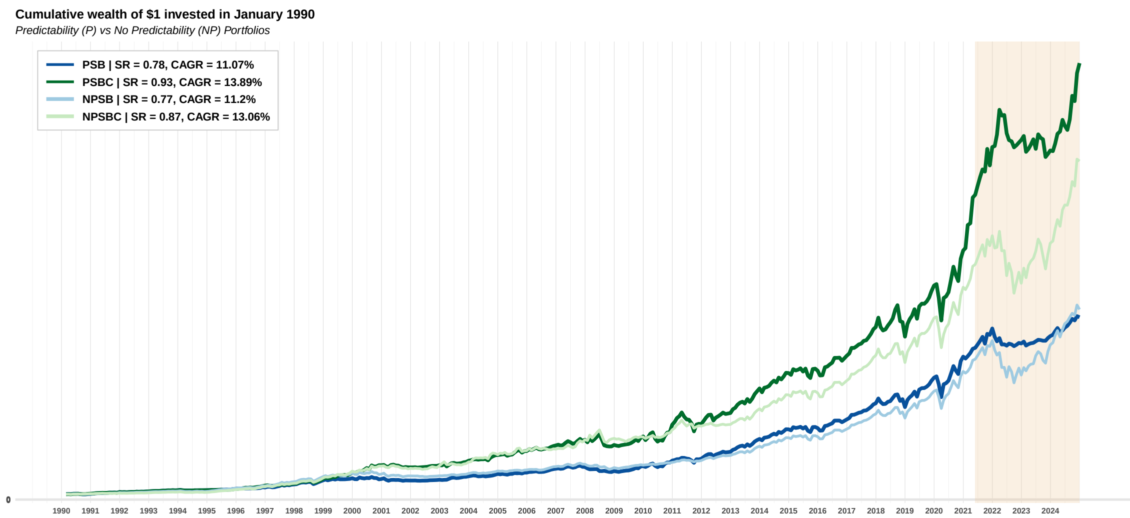

Figure A.1: Wealth paths for different alternatives. *(Extended Sample)*

References

Ang, A. and Bekaert, G. (2007). Stock Return Predictability: Is it There? *The Review of Financial Studies*, 20(3):651–707.

Anson, M. J. P. (1998). Spot Returns, Roll Yield, and Diversification with Commodity Futures. *Journal of Alternative Investments*, 1(3):16–32.

Avramov, D. and Chordia, T. (2006). Predicting stock returns. *Journal of Financial Economics*, 82(2):387–415.

Bampinas, G. and Panagiotidis, T. (2015). Are gold and silver a hedge against inflation? A two century perspective. *International Review of Financial Analysis*, 41:267–276.

Basak, S. and Pavlova, A. (2016). A model of financialization of commodities. *The Journal of Finance*, 71(4):1511–1556.

Baur, D. G. and McDermott, T. K. (2010). Is gold a safe haven? International evidence. *Journal of Banking & Finance*, 34(8):1886–1898.

Bekkers, N., Doeswijk, R. Q., and Lam, T. (2009). Strategic Asset Allocation: Determining the Optimal Portfolio with Ten Asset Classes. *Journal of Asset Management*. Available at SSRN 1368689.

Bessler, W. and Wolff, D. (2015). Do commodities add value in multi-asset portfolios? An out-of-sample analysis for different investment strategies. *Journal of Banking & Finance*, 60:1–20.

Bodie, Z. and Rosansky, V. I. (1980). Risk and Return in Commodity Futures. *Financial Analysts Journal*, 36(3):27–39.

Brixton, A., Brooks, J., Hecht, P., Ilmanen, A., Maloney, T., and McQuinn, N. (2023). A Changing Stock-Bond correlation: Drivers and Implications. *Journal of Portfolio Management*, 49(4):1–15.

Büyükşahin, B. and Robe, M. A. (2014). Speculators, commodities and cross-market linkages. *Journal of International Money and Finance*, 42:38–70.

Campbell, J. Y. and Shiller, R. J. (1988). The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. *The Review of Financial Studies*, 1(3):195–228.

Campbell, J. Y. and Thompson, S. B. (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *The Review of Financial Studies*, 21(4):1509–1531.

Chen, N.-F., Roll, R., and Ross, S. A. (1986). Economic Forces and the Stock Market. *Journal of Business*, 59(3):383–403.

Cochrane, J. H. (2008). The Dog That Did Not Bark: A Defense of Return Predictability. *The Review of Financial Studies*, 21(4):1533–1575.

Commodity Futures Trading Commission (2008). Staff report on commodity swap dealers and index traders with Commission recommendations. *CFTC Staff Report*.

Daskalaki, C. and Skiadopoulos, G. (2011). Should investors include commodities in their portfolios after all? New evidence. *Journal of Banking & Finance*, 35(10):2606–2626.

Daskalaki, C., Skiadopoulos, G., and Topaloglou, N. (2017). Diversification benefits of commodities: A stochastic dominance efficiency approach. *Journal of Empirical Finance*, 44:250–269.

Domanski, D. and Heath, A. (2007). Financial investors and commodity markets. *BIS Quarterly Review*.

Dunsby, A. and Nelson, K. (2010). A Brief History Of Commodities Indexes. *Journal of Indexes*, 13(3):36–39.

Estrella, A. and Hardouvelis, G. A. (1991). The Term Structure as a Predictor of Real Economic Activity. *The Journal of Finance*, 46(2):555–576.

Fama, E. F. (1990). Term-structure forecasts of interest rates, inflation and real returns. *Journal of Monetary Economics*, 25(1):59–76.

Fama, E. F. and French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1):3–25.

Fama, E. F. and French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, 25(1):23–49.

Fama, E. F. and Schwert, G. W. (1977). Asset returns and inflation. *Journal of Financial Economics*, 5(2):115–146.

Fethke, T. and Prokopczuk, M. (2018). Is commodity index investing profitable? Technical report, Hannover Economic Papers (HEP).

Goyal, A. and Welch, I. (2003). Predicting the Equity Premium with Dividend Ratios. *Management Science*, 49(5):639–654.

Goyal, A. and Welch, I. (2008). A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. *The Review of Financial Studies*, 21(4):1455–1508.

Goyal, A., Welch, I., and Zafirov, A. (2024). A Comprehensive 2022 Look at the Empirical Performance of Equity Premium Prediction. *The Review of Financial Studies*, 37(11):3490–3557.

Hull, B., Qiao, X., and Bakosova, P. (2017). Return Predictability and Market-Timing: A One-Month Model. *Journal of Investment Management*, 17(2):47–64.

Hull, J. C. (2014). *Options, Futures, and Other Derivatives*. Pearson Education, 9th edition.

Jiang, H., Skoulakis, G., and Xue, J. (2018). Oil and Equity Return Predictability: The Importance of Dissecting Oil Price Changes. *Robert H. Smith School Research Paper*, 2822061.

Jobson, J. D. and Korkie, B. M. (1981). Performance Hypothesis Testing with the Sharpe and Treynor Measures. *Journal of Finance*, 36(4):889–908.

Keim, D. B. and Stambaugh, R. F. (1986). Predicting Returns in the Stock and Bond Markets Predicting Returns in the Stock and Bond Markets. *Journal of Financial Economics*, 17(2):357–390.

Kremer, P. (2015). Comparing Three Generations of Commodity Indices: New Evidence for Portfolio Diversification. *Alternative Investment Analyst Review*, 1.

Ledoit, O. and Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, 15(5):850–859.

Miffre, J. (2012). Comparing First, Second, and Third Generation Commodity Indices. *EDHEC-Risk Institute Working Paper*.

Molenaar, R., Sénéchal, E., Swinkels, L., and Wang, Z. (2024). Empirical Evidence on the Stock–Bond Correlation. *Financial Analysts Journal*, 80(3):17–36.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Rallis, G., Miffre, J., and Fuertes, A.-M. (2013). Strategic and Tactical Roles of Enhanced–Commodity Indices. *Journal of Futures Markets*, 33(10):965–992.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2013). International Stock Return Predictability: What Is the Role of the United States? *The Journal of Finance*, 68(4):1633–1662.

Ruano, F. and Barros, V. (2022). Commodities and portfolio diversification: Myth or fact? *The Quarterly Review of Economics and Finance*, 86:281–295.

Silvennoinen, A. and Thorp, S. (2013). Financialization, crisis and commodity correlation dynamics. *Journal of International Financial Markets, Institutions and Money*, 24:42–65.

Štulec, I., Baković, T., and Džalto, I. (2024). Commodities and portfolio diversification: Empirical evidence from crisis duress. *InterEULawEast: Journal for the International and European Law, Economics and Market Integrations*, 11(1):73–100.

Wang, Y., Hao, X., and Wu, C. (2021). Forecasting stock returns: A time-dependent weighted least squares approach. *Journal of Financial Markets*, 53:100568.

Yan, L. and Garcia, P. (2017). Portfolio Investment: Are Commodities Useful? *Journal of Commodity Markets*, 8:43–55.