UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Predicting Global Stock Returns Using Commodities: A Gradient Boosting Decision Tree Approach

Jesper Frisell
Philip Lindstedt

Supervisor: Jonas Frey
Master’s thesis in Finance
Spring 2025
Graduate School, School of Business, Economics and Law
University of Gothenburg, Sweden

Abstract

We examine the predictability of stock returns using commodity futures prices across 39 countries from 1999 to 2024 using the XGBoost implementation of the Gradient Boosting Decision Tree approach. There is evidence of increased integration between commodities and stock markets. Despite this, research examining the predictability of commodities on stock returns is limited, especially on a global scale. The aim is to build on previous studies and explore heterogenous effects of commodity price changes on countries. We find evidence of predictability for four individual commodities and two commodity indices after sampling twelve commodities and four indices. Copper and crude oil show the strongest predictability among individual commodities, while industrial metals and energy demonstrate the strongest predictability among commodity indices. Our results also indicate strong heterogeneous effects, with some countries exhibiting significantly greater exposure to commodity prices. In particular, the Australian stock market is more exposed to price changes in copper and industrial metals, while the Norwegian market shows large sensitivity to oil and energy. Based on our findings, a competitive long-short trading strategy is proposed.

Keywords: Gradient Boosting Decision Tree; XGBoost; Stock return predictability; Feature importance; SHAP values; Pooling

Acknowledgements

We sincerely thank our supervisor, Jonas Frey, for his continued support. His valuable insights and guidance were instrumental in writing this thesis.

Contents

1. Introduction

We investigate the predictability of commodity returns on stock returns using a Gradient Boosting Decision Tree (GBDT) approach. Heterogeneous effects are also explored by isolating the impact of commodity price changes at the country level, building on previous research by Black et al. (2014), Jordan et al. (2016), Salisu et al. (2019, Jacobsen et al. (2019), Iyke and Ho (2021) and Fasanya et al. (2023). By analyzing in-sample, out-of-sample, and pooled predictions, we provide insights of the predictive properties of each commodity on the representative stock market of each country in our sample. A trading strategy is then proposed based on these predictive relationships. Our main findings indicate that copper, crude oil, and industrial metals, strongly predict global stock prices, while gold and silver exhibit poor predictive power.

The contribution of this paper is threefold. First, we use a more diverse and comprehensive dataset than has previously been done by sampling the stock markets of 39 countries, with the most globally representative out of earlier papers on commodity-stock predictability sampling 10 countries. This allows us to observe geographic heterogeneity in the effects of commodity price changes to a greater extent, where stock markets in diverse countries and regions are impacted differently. By using daily data, our dataset also includes more data points, with previous work using weekly, monthly, or quarterly sampling.

Second, by providing a recent dataset ranging from 1999 to 2024, our data extends beyond prior studies ending in 2018. This is significant considering the trend towards increased commodity financialization. Previous papers also do not cover the two latest major global crises, which are the COVID-19 crisis starting in 2020 along with the Russian invasion of Ukraine in 2022. Extreme events are shown to impact the interconnectedness between commodities and stock prices, making this an important contribution (Dai & Zhu, 2022; Mensi et al., 2024; Naeem et al., 2024).

Third, while previous studies utilize various econometric techniques, we apply machine learning methods by employing a GBDT model. Machine learning has been proven to outperform in some cases through better modeling of non-linear and heteroscedastic relationships, which are often present in stock price prediction (Gu et al., 2020).

Stock price prediction remains one of the most widely covered areas of economic research. The commonly referenced early papers by Fama (1965, 1970) propose the Efficient Market Hypothesis (EMH), arguing that stock prices reflect historical data along with public available information, supporting the weak and semi-strong forms of market efficiency. Since then, several studies provide international evidence of predictors such as interest rates, dividend yields, and US lagged returns, which can be used to predict stock prices, thereby challenging the semi-strong form of market efficiency (Ang & Bekaert, 2007; Campbell, 1987; Campbell & Shiller, 1988; Fama & French 1988, 1989; Hjalmarsson, 2010; Lewellen, 2004; Rapach et al., 2013; Solnik, 1993). With predictable components now well established, economic research has increasingly focused on identifying robust predictors and explanatory variables.

The complexity of stock price prediction subjects it to a large body of research that seeks to find suitable forecasting methods to detect these predictors and yield superior returns for investors (Thakkar & Chaudhari, 2021). Early methods include the autoregressive moving average, and autoregressive integrated moving average, which focus on the time-series dimension. These techniques struggle, due to the non-linear and complex nature of stock price prediction, leading to new methods such as the Generalized Autoregressive Conditional Heteroskedasticity model. Advancements in computational technology drive the development of machine learning, which show effectiveness in analyzing the cross-sectional dimension of data. These machine learning techniques include decision trees and support vector machines. Further techniques, such as Random forest and GBDT models, have since been developed and used for stock price prediction (Cakici et al., 2023).

There is a trend towards increased integration between the commodity and stock markets, with investors increasingly purchasing commodities for portfolio management purposes, rather than physical delivery, a trend commonly known as “commodity financialization” (Ali et al., 2020; Tang & Wei, 2012). Researchers have extensively studied the interconnectedness between the two markets, which has been found to strengthen during major global crisis events, such as the Global Financial Crisis, the European Debt Crisis, and the COVID-19 pandemic (Dai & Zhu, 2022; Mensi et al., 2024; Naeem et al., 2024). The most recent and still ongoing crisis, started in February 2022, with the Russian invasion of Ukraine, directly impacts the commodity markets due to its nature. Russia is one of the world’s major exporters of crude oil, natural gas, gold, and wheat, while Ukraine is a major supplier of wheat and corn (OEC, 2023a; OEC, 2023b). As a result of the conflict and imposed sanctions, commodity prices changed drastically, having varying effects across countries (Jacks & Stuermer, 2020; Aizenman et al., 2024; OECD, 2022).

Although the increased integration between commodities and the stock market is well documented, global evidence on commodities as a predictor remains scarce, compared to predictors such as interest rates and dividend yields. For the main literature, we found six papers that studies a set of commodities as predictors. Black et al. (2014) examine the US stock market, finding conflicting results where predictability depends on the sample period. Jordan et al. (2016) sample the Canadian market finding mixed results across sectors and commodities. Salisu et al. (2019) explore G7 countries and concludes that commodities have predictive power for most countries. Jacobsen et al. (2019) sample 10 countries, providing evidence that mainly industrial metal commodities to have predictive powers for most countries. Iyke and Ho (2021) study the Netherlands, the UK, and the US, finding predictive powers for many commodities, although mainly in-sample. The most recent paper by Fasanya et al. (2023) focuses on BRICS countries and finds predictability for most sampled commodities for Brazil, Russia, and South Africa. We aim to contribute to the literature and answer the following three research questions:

> *RQ1: Can a Gradient Boosting Decision Tree model find predictability for commodities on stock returns?*
>
> *RQ2: How do commodity price changes affect stock returns differently across countries?*
>
> *RQ3: Can a trading strategy be developed using commodities to predict stock returns*?

2. Literature Review

The following section presents relevant literature, starting with the most central research where the predictability of a wide range of commodities in various stock markets is examined. More niche papers are then brought up that study the predictability of a single commodity, or discusses co-movement between commodities and stock markets, rather than predictability. Afterward, established research on common predictors is specified, followed by a section on the performance of trading strategies using these common predictors.

2.1 Main literature

Black et al. (2014) study the effect of the S&P GSCI Commodity Index on the US stock market using quarterly data ranging from 1973 to 2012. Their methodology includes a cointegration analysis, predictive regressions, and rolling forecasts. Two structural breakpoints are identified in 1994 and 2002, marking the start and end of the dot-com crash, where the relationship between commodities and stocks changes significantly. Before 1994, commodity and stock prices exhibit a negative predictive relationship, where an increase in commodity prices is associated with a decrease in stock prices. However, the authors find that the predictive relationship becomes neutral between 1994 and 2002 and turns positive after 2002.

Jordan et al. (2016) investigate whether commodity returns can predict stock returns for eight Canadian equity sectors. They use the bootstrap aggregating method, alongside forecast combination methods, to compare with a benchmark AR (1) model. For their sample, they use weekly data ranging from 1985 to 2011, using 23 commodities including cattle, coffee, corn, crude oil, gas, heating oil, silver, sugar, and wheat. Their findings are that predictability is mixed across both sectors and predictor commodities, with only gold and silver being consistent in their predictability of stock returns out-of-sample.

Salisu et al. (2019) evaluate the predictive power of commodity prices on stock returns in G7 countries (Canada, France, Germany, Italy, Japan, UK, US). They use an extended model to account for macroeconomic risks, and address issues of asymmetry, conditional heteroscedasticity, endogeneity, persistence, and structural breaks. Their sample consists of monthly data covering 1960 to 2017 for some commodities, using agriculture, energy, oil, industrial metals, precious metals, and a non-energy commodity index as commodity features. While the predictability of individual commodities is not highlighted, their results support that commodities have predictive power, as their model outperforms the historical mean model, both in-sample and out-of-sample, except for Japan. The result for the US varies depending on the amount of data used, where a larger dataset makes the historical model superior.

Jacobsen et al. (2019) focus on metals using the S&P GSCI Industrial Metals Index, with additional attention on copper and aluminum, the primary metals within the index. They use stock indices for 10 modern countries: Australia, Canada, France, Germany, Italy, Japan, Netherlands. Sweden, Switzerland, and UK using monthly data ranging from 1977 to 2016. By employing a state-switching regression model, the authors find that a price increase of metals predicts a decrease in stock prices during market expansions, and an increase in recessions. This result is statistically significant for four countries during expansions, and nine countries in recessions, with a monthly out-of-sample R^2 of three percent to eight percent. Although the study also includes energy, precious metals, livestock, and agriculture indices in their sample, no significant state-switching predictability is found for these.

Iyke and Ho (2021) study commodity predictability in the markets of the Netherlands, UK, and US. They apply a predictive regression using a biased adjusted Generalized Least Squares estimator, an Augmented Dickey-Fuller test, an AR (1) coefficient, and an endogeneity test. For their sample, they use monthly data on 25 individual commodities, including copper, coffee, cotton, gold, oil, silver, sugar, and wheat, as well as the four indices of agriculture, energy, livestock, and metals, with a maximum sample period from 1629 to 2005 for some commodity features. Both sub-samples and the complete sample period are examined, and significant in-sample predictability is found for most commodities including beef, coffee, cotton, silver, sugar, and wheat, as well as all of the indices. Out-of-sample results are less widespread, with significant predictability found in both the 1871 to 2005 and 1629 to 2005 intervals for coal, lead, tea, zinc, and metals.

Fasanya et al. (2023) explore predictability on the BRICS countries (Brazil, Russia, India, China, and South Africa). A Feasible Quasi Generalized Least Squares approach is used as the predictive model to address issues of heteroskedasticity, serial dependence, persistence, and endogeneity. Their sample consists of 10 commodities including beverages, grains, oil, metals as well as precious metals using monthly data from 1994 to 2018. Significant out-of-sample predictability is found for Brazil, Russia, and South Africa for virtually all commodities, except for India and China. The authors also highlight the importance of asymmetries, where positive and negative commodity price changes can impact stock prices differently, supported through a Campbell-Thompson test and root-mean-square error (RMSE) comparisons.

2.2 Research on individual commodities

Previous studies examine the relationship between individual commodities—or commodity indices—and with stock markets. These studies are narrower in scope compared to the four primary studies previously mentioned, as they are limited to studying a particular commodity or country, and do not discuss predictability, but rather interconnectedness and co-movement.

Copper is sometimes referred to as “Doctor Copper” due to its predictive ability of the global economy (Chen, 2022). In line with this, several studies support copper as a predictor of international stock markets (Jacobsen et al., 2019; Fasanya et al., 2023).

Previous studies on the relationship between oil and the stock market are contradictory, with some studies showing that it influences stock prices. However, the direction of the effect differs between studies and samples (Arfaoui & Rejeb, 2017). Some research supports that oil is a net transmitter of shocks, where a price shock in oil, is followed by a spillover effect on the stock market (Liao et al., 2021; Wang et al., 2019). However, some papers find that, depending on the market state, oil can instead be a net receiver, where the stock price affects oil prices (Biswas et al., 2024; Mensi et al., 2024). A positive correlation of crude oil and other commodities, has been identified and found to have increased dramatically since before the financial crisis in 2008, with the correlation between crude oil and copper being particularly strong (U.S. Energy Information Administration, 2025).

One study finds a significantly weaker co-movement with stock prices for natural gas than for oil (Mensi et al., 2021). Additional work supports the idea that while there is a strong interconnectedness between natural gas and stock prices, natural gas tends to be a net receiver of volatility (Geng et al., 2021; Gong et al., 2021).

Wheat is shown to exhibit significant spillover effects with stock markets, typically acting as a receiver of spillovers, rather than a transmitter (Garcia-Jorcano & Sanchis-Marco, 2022). A recent paper finds that wheat takes on the role of a net transmitter during the Russia-Ukraine war, contrary to its usual state of being a net receiver (Biswas et al., 2024).

Numerous studies support gold as a hedging tool and its legitimacy as a safe haven due to its negative correlation to the stock market (e.g., Baur & Dermott, 2010; Shahzad et al., 2020). Most research also shows that gold is a net receiver of shocks (Dai et al., 2022; Mensi et al., 2024).

2.3 Established predictors

Established papers studying out-of-sample predictability of stock returns identify several significant predictor variables. Campbell and Thompson (2008) analyze monthly US stock market data ranging from 1927 to 2004 using 15 forecast variables including the divided-price ratio, earnings-price ratio, term spread, book-to-market ratio, and inflation. They demonstrate that even small positive R^2 statistics are relevant for investors, as they can significantly improve portfolio performance.

Hjalmarsson (2010) examines four standard forecasting variables, which are the earning-price ratio, the dividend-price ratio, the short-term interest rate and the term spread, and finds out-of-sample R^2 values ranging from -1.881 percent to 1.939 percent. In-sample R^2 and pooled R^2 results varied from 0.001 percent to 1.973 percent in-sample, and -1.898 percent and 2.970 percent for the pooled regression. The results indicate that the short-term interest rate and the term spread are robust predictors of stock returns in developed markets.

Rapach et al. (2013) examine stock return predictability in 11 industrialized countries, focusing on the influence of US stock returns. Analyzing monthly data from 1985 to 2010, the study finds that lagged US returns significantly forecast returns in non-U.S. stock markets. Out-of-sample R^2 values range from -0.69 percent to 3.81 percent, with pooled R^2 scores between 0.5 percent and 3.88 percent. This predictability translates into utility gain for investors.

2.4 Previous Trading Research

There is no existing established study that constructs a trading strategy using commodity prices to predict returns in a market-neutral long-short setting. However, while no previous literature examines an identical setting, numerous market-neutral strategies exist that use other predictors. Examples include work by Gatev et al. (2006), Asness et al. (2013), Blitz et al. (2019), and Lin et al. (2021), who evaluate market-neutral trading strategies based on pair-trading, combinations of value and momentum, volatility, and pattern recognition, respectively. Their strategies achieve Sharpe ratios ranging from 0.59 to 1.45. This wide range is partly explained by different approaches to transaction costs and risk-free rates, with the lower reported Sharpe ratios using more conservative values. Asness et al. (2013) achieve the highest Sharpe ratio out of the mentioned strategies when ignoring transaction costs and using gross returns. The value of risk-free rate is not always explicitly stated, but can be assumed to be different due to the differing periods of the studies, with short-term rates typically being used.

3. Theory

This section reviews central theoretical concepts by outlining the fundamental principles of machine learning and economic theory relevant to the study.

3.1. Machine Learning

Machine learning is a subfield of artificial intelligence that tries to make as accurate predictions as possible of an outcome variable by learning the relationship between the outcome and the predictor variables from training data. Data is the building block of machine learning, which can be stock prices, returns, or other variables. Generally, the larger the dataset, the better the model's performance. After gathering data, the practitioner selects an appropriate model, enabling the computer to learn by locating patterns or predicting outcomes. The parameters are then adjusted and tuned to improve the model's accuracy. One method to evaluate the model's accuracy is to withhold a proportion of the data during training and then measure the performance compared to the unseen data. Machine learning has three subcategories: supervised, unsupervised, and reinforcement. This study applies supervised machine learning, which trains the model using labeled datasets, allowing it to identify patterns and improve accuracy over time. (Brown, 2021).

3.1.1 Why apply Machine Learning?

A study by Gu et al. (2020) finds that in high-dimensional financial datasets, particularly when the number of predictors exceeds the number of observations, machine learning methods, such as tree-based models, outperform leading regression-based strategies in performance. While econometric models focus on testing the statistical significance of variables on an outcome, machine learning focuses on predicting the outcome as accurately as possible. The authors praise machine learning tools for their application in portfolio optimization, demonstrating superior performance compared to leading regression-based strategies and achieving higher Sharpe ratios. They also state that machine learning is suitable when working with macroeconomic variables and financial data due to the high correlation of the predictors.

3.1.2 Limitations & Risks

While machine learning comes with many opportunities and flexibility over traditional econometric prediction techniques, it brings a risk of overfitting, which is its primary issue.

Overfitting in supervised machine learning causes the model to fail to generalize from the observed data to new, unseen data. Figuring out why overfitting occurs can be a complex problem to resolve. Possible causes include noisy data, a training set that is either too small or contains irrelevant information, and too high complexity. Characteristics of an overfitting model is strong performance on the training (in-sample) set, but weak performance on the validation (out-of-sample) set (Ying, 2019).

3.1.3 Gradient Boosting Decision Trees

A decision tree model is a machine learning algorithm that splits data into branches based on feature values, forming a tree-like structure, where each node represents a decision rule. The Gradient Boosting Decision Tree (GBDT) algorithm combines decision trees with boosting, where multiple weak decision trees are trained sequentially to form a strong model. Gradient boosting extends the idea of boosting by setting targeted outcomes for the next model to minimize errors. In a GBDT model, the process begins by constructing an initial decision tree. Subsequent trees are added sequentially, each learning from the residuals of the previous tree. The model learns by optimizing a given loss function through gradient descent to improve its predictive accuracy gradually. Since its introduction, new implementations of the GBDT model, such as XGBoost, have been developed. XGBoost is known for its dominating performance in machine learning competitions, such as the Kaggle due to its speed and treatment of overfitting. It reduces overfitting through regularization, shrinkage, subsampling, and early stopping. (Chen & Guestrin, 2016; Friedman, 2001; Machine learning plus, 2021; NVIDIA, 2025; Zhou et al., 2019).

Table 1: Data for a simplified example to illustrate how a GBDT model works. Panel A shows an overview of the initial training data given to the model, which in this case is the number of hours studied, hours slept, and exam scores. Panel B shows the predicted exam scores based on the training data. The exam score of student 5 is unknown and must be predicted based on the training data. The rightmost column displays the residuals (which is the difference between actual and predicted scores) based on which the model will seek to improve on its next prediction.

Student	Hours Studied (x_1)	Hours Slept (x_2)	Exam Score (y)	Predicted (y)	Residual (error)
Panel A				Panel B	
1	1	4	40	45	-5
2	2	5	50	45	5
3	3	7	70	70	0
4	4	7	80	80	0
5	5	4	?	75	

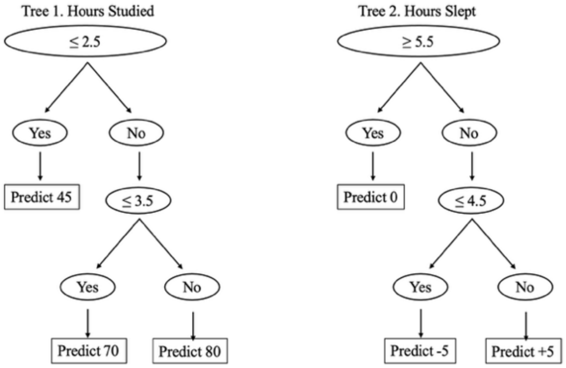


Figure 1: Two decision trees are illustrated. Tree 1 uses the hours studied variable, first splitting at ≤2.5 hours (predicting 45 if yes) and then again at ≤3.5 hours (predicting 70 if yes and 80 if no). Tree 2 uses the hours slept variable to improve on the residuals (erroneous predictions) of Tree 1. Tree 2 first splits at ≥5.5 hours (predicting 0 if yes) and then again at ≤4.5 hours (predicting -5 if yes, and +5 if no).

Table 1 illustrates a simplified example of how decision trees work. In this example, decision trees predict exam scores based on the two features of hours studied and hours slept. For students 1 to 4, the exam scores are known and they therefore make up the training data for the model. The objective is to predict the exam score of student 5, which is unknown and constitutes the validation set in this example. Panel A of Table 1 provides an overview of the sample.

The model constructs the first decision tree (Tree 1) in Figure 1 using the training data from Panel A of Table 1, which contains the data for students 1 to 4. Predicted exam scores at each leaf (which is the terminal node where the final prediction is made) are estimated by averaging the scores of the students who fall into that category based on similar hours studied. A second tree is then constructed based on the residual errors from Tree 1, calculated as the differences between the actual exam scores and the predicted scores from Tree 1. The first tree correctly predicted the exam scores of students 3 and 4, but not for students 1 and 2, as shown in Panel B of Table 1.

The model then constructs a second tree (Tree 2) to improve the predictive performance of the first tree, by using another feature variable, which in this case is hours slept. Tree 2 uses the hours slept variable to fit the training data better, making the predictions accurate for all students. When predicting the exam score for student 5, Tree 1 predicts a score of 80. Tree 2, which tried to improve the first, instead predicts a score of 75 for student 5 by adding a penalty for low hours of sleep. In practice, the model does not necessarily test features separately but rather tunes many parameters simultaneously and construct several hundred —or even thousands of trees —to improve fit to the training set and create a strong model to predict the validation set. Still, this example illustrates how trees make decisions to improve on previous trees based on prediction errors.

3.1.4 Black Box in Machine Learning

The high complexity of machine learning makes interpreting their results more challenging compared to using traditional regression models. The concept of a "black box" is frequently brought up in literature, and researchers have proposed different solutions to this problem. Several methodologies have been developed such as Shapley Additive explanation (SHAP), Local Interpretable Model-agnostic Explanations, Deeplift, Partial Dependence Plots and the gain method being used to illustrate and study the output. While these tools do not fully open the box, studies generally find that interpretability has improved, with newer methodologies such as SHAP values outperforming more traditional approaches (Boehmke & Greenwell, 2019; Carmona et al, 2022; García & Aznarte, 2020; Lundberg et al., 2020).

3.2 Efficient Market Hypothesis

The EMH, developed by Eugene Fama (1965, 1970), posits that financial markets efficiently incorporate all available information, making it impossible to achieve above-average returns consistently through analysis or market timing. In his 1965 paper, Fama provides supporting evidence that stock prices follow a random walk, meaning that past price movements cannot predict future prices. The concept of a random walk implies that all known information is already reflected in current prices, making technical analysis ineffective. Fama further refines the EMH in his 1970 paper by categorizing market efficiency into three forms. Weak-form market efficiency means historical prices cannot predict future stock movement. Semi-strong form efficiency asserts that all publicly available information is already incorporated into stock prices and, therefore, cannot be used to achieve abnormal returns through forecasting. Strong-form efficiency extends this concept further by stating that even private, insider information is fully reflected in stock prices. Both weak-form and semi-strong-form efficiency is supported in Fama's early papers. However, the semi-strong form has since been challenged by many researchers, including Fama himself (Ang & Bekaert, 2007; Campbell, 1987; Campbell & Shiller, 1988; Fama & French 1988, 1989; Hjalmarsson, 2010; Lewellen, 2004; Rapach et al., 2013; Solnik, 1993). Still, the framework by Fama is a cornerstone of economic research and a fundamental concept in the context of stock price prediction.

3.3 Modern Portfolio Theory and Sharpe Ratio

Harry Markowitz (1952) laid the foundation for portfolio management by introducing Modern Portfolio Theory (MPT) in his seminal paper. MPT emphasizes that investors should optimize portfolios by maximizing expected return, while minimizing variance, and considering covariances between securities, this promotes diversification and risk minimization. Building on the MPT, William F. Sharpe (1964, 1966) extends the theory in his papers by developing the Capital Asset Pricing Model, introducing the concepts of systematic- and unsystematic risk as well as establishing a linear relationship between an asset's expected return and its market risk. Sharpe further advanced performance evaluation in his 1966 paper, where he introduced the reward-to-variability ratio, now commonly known as the Sharpe ratio. The Sharpe ratio has become a widely used tool for portfolio performance evaluation, where the portfolio with the highest Sharpe ratio is the best portfolio from a risk-return trade-off perspective.

4. Data and Methodology

The following section begins by describing the dataset used in the analysis, followed by an outline of the methodology, and concludes with the trading strategy setup.

4.1 Data Collection

We collect daily closing prices from 1999 to 2024 for the most representative stock indices of 39 countries, using data from Standard and Poor's Compustat Global via Wharton Research Data Services (Standard & Poor's, n.d.). This period is selected to maximize the number of indices with complete and consistent data, ensuring robust cross-country comparison. Starting from 1999, this provides enough data while also covering major global crisis events, such as the dot-com bubble burst in 2000 and the Global Financial Crisis in 2007. We convert the prices of stock indices, commodity futures, and commodity indices to log returns. A complete list of the countries studied, and their representative stock index is given in Table A1.

For commodities, we collect daily futures closing prices, from 1999 to 2024 for 12 individual commodities: cattle, coffee, copper, corn, cotton, crude oil, gold, heating oil, natural gas, silver, sugar, and wheat. Data is also collected for four major commodity index categories: agriculture, energy, industrial metal, and precious metal. The data for individual commodities and commodity indices is collected from Investing.com (n.d.) and S&P Capital IQ (n.d.) respectively. Table A2 shows a complete list of the commodities in the sample.

4.2 Methodology

The methodology section is divided into subsections to describe the model framework, commodity predictability evaluation metrics, and feature importance analysis used in the study.

4.2.1 Model Specification

We apply the XGBoost implementation of the GBDT to predict the return of stock indices, following the approach proposed by Chen and Guestrin (2016). To understand this approach, consider a dataset $D = \{(x_i, y_i)\}_{i=1}^{n}$ where each feature vector $x_i \in \mathbb{R}^m$ contains commodity futures and indices log returns, and where the corresponding target $y_i \in \mathbb{R}$ represents stock index log returns. The model uses K additive functions to predict returns, as shown in equation (1):

(1) $$\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i), \qquad f_k \in \mathrm{F},$$

where x_i is the feature vector for the i-th observation, K is the number of trees in the model, $f_k(x_i)$ is the k$^{\text{th}}$ tree's output, and F represents the space of all possible regression trees. Each tree f_k, maps x_i to a leaf node containing a weight value. The objective function is to minimize prediction error and regularization terms:

(2) $$\mathcal{L}(\phi) = \sum_{i} l(y_i, \hat{y}_i) + \sum_{k} \Omega(f_k),$$

where the regularization term is defined as:

(3) $$\Omega(f) = \gamma T + \tfrac{1}{2}\lambda \|w\|^2.$$

In equation (2), l denotes the loss function where the squared error is $l = (y_i - \hat{y}_i)^2$. The first term represents the sum of squared errors across all observations, while the second term acts as a regularization component. This regularization discourages overfitting of the training data by penalizing overly complex trees that may otherwise capture noise in commodity- or index returns. At iteration t, the model adds a new tree f_t to improve predictions; formally $\hat{y}_i^t = \hat{y}_i^{(t-1)} + f_t(x_i)$, the objective function to minimize at this step becomes:

(4) $$\mathcal{L}^t = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t).$$

Using a second-order Taylor expansion of the loss function based on the principle of mean forward stagewise additive modeling:

$$\mathcal{L}^t \simeq \sum_{i=1}^{n}\left[l\left(y_i,\hat{y}_i^{(t-1)}\right)+g_i f_i(x_i)+\frac{1}{2}h_i f_t^2(x_i)\right]+\Omega(f_t), \tag{5}$$

where the first derivative $g_i = \partial_{\hat{y}^{(t-1)}} l\left(y_i,\hat{y}_i^{(t-1)}\right)$ and the second derivative are $h_i = \partial^2_{\hat{y}^{(t-1)}} l\left(y_i,\hat{y}_i^{(t-1)}\right)$. Since $l\left(y_i,\hat{y}_i^{(t-1)}\right)$ is constant with respect to f_t, it can be dropped, an approximated objective function is derived:

$$\tilde{\mathcal{L}}^t = \sum_{i=1}^{n}\left[g_i f_t(x_i)+\frac{1}{2}h_i f_t^2(x_i)\right]+\Omega(f_t). \tag{6}$$

The model further represents the relationship between branch structures and their corresponding leaf weights using the following expression:

$$f_k(x) = w_{q(x)},$$

where $q(x)$ denotes the leaf index to which the input sample x is assigned and $w \in \mathbb{R}^T$ represents the set of leaf weights as a t-dimensional real-valued vector. The decision tree's complexity depends on the number of leaves T and the L_2-norm of all weight vectors. Thus, writing the regularization term in equation (3) as:

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2 \,. \tag{7}$$

Let $I_j = \{i[q(x_i)\}) j$ denote the set of samples assigned to the leaf node j. For a fixed tree structure $q(x)$, the optimal weight w_j^* of leaf j is given by:

$$w_j^* = -\frac{\sum_{i\,\in\, I_j} g_i}{\sum_{i\,\in\, I_j} h_i + \lambda}, \tag{8}$$

where g_i and h_i are the first- and second-order gradient statistics, respectively. The objective function can thus be reformulated as:

$$(9) \qquad \tilde{\mathcal{L}}^t(q) = -\frac{1}{2}\sum_{j=1}^{T}\frac{\left(\sum_{i \in I_j} g_i\right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T.$$

Where equation (9) serves as a scoring function to assess the quality of a tree structure (q). This score is similar to the impurity score used for evaluation decision trees, but it is adapted for a broader range of objective functions. Normally, $\tilde{\mathcal{L}}^t(q)$ evaluates the loss of the entire tree by summing the contribution of all leaves and lists all possible tree structures of (q) but, it is computationally infeasible, since the number of possible trees grows exponentially with the number of nodes or features. To solve this, the model uses a greedy algorithm that begins with a single leaf and adds branches. The model decides to split at each node by evaluating the potential loss reduction. Assuming I_L and I_R are the instance sets of the left and right nodes post-split, with $I = I_L \cup I_R$, the model calculates the loss reduction after the split as:

$$(10) \qquad \mathcal{L}_{split} = \frac{1}{2}\left[\frac{\left(\sum_{i \in I_L} g_i\right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i\right)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{\left(\sum_{i \in I} g_i\right)^2}{\sum_{i \in I} h_i + \lambda}\right] - \gamma.$$

4.2.2 Expanding Window Approach

An expanding window approach is used by initially training the model on the first 10 years of data (from 1999 to 2008) and then validating it in the following year (2009). With each iteration, the training window expands to include one additional year of data, and the subsequent year is used for validation. This approach ensures that the current period is excluded from the validation sample to prevent look-ahead bias. The length of the initial training period is selected by minimizing the out-of-sample RMSE to prevent underfitting and overfitting, as shown in Figure A1. Our approach aligns with the expanding methodologies discussed by Stock and Watson (2007).

4.2.3 Hyperparameter tuning

Hyperparameter tuning is a crucial step in fitting a machine learning algorithm, as poorly tuned parameters can reduce model performance and increase the risk of overfitting. Seven parameters are used and tuned based on best practices: n_estimators, learning_rate, max_depth, subsample, colsample_bytree, alpha, and lambda. N_estimators define the number of trees, tuned to balance model complexity with overfitting risk, with more trees improving the fit but increasing computation time. Learning_rate shrinks each tree's contribution, adjusting to ensure gradual learning. Lower values enhance accuracy but require more trees. Max_depth limits tree depth, tuned to capture complex patterns without overfitting by restricting how much each tree can split. Subsample sets the fraction of data used per tree to introduce randomness and reduce overfitting on noisy data. Colsample_bytree controls the fraction of features sampled per tree, adjusted to prevent reliance on dominant features and improve generalization. Alpha (L1 regularization) encourages sparsity in weights, tuned to simplify the model and enhance interpretability when relevant features vary. Lambda (L2 regularization) penalizes large weights, helping to smooth predictions and reduce overfitting in high-variance scenarios (Boehmke and Greenwell, 2019; De Prado, 2018).

Cross-validation is performed using the GridSearchCV function in Python. Splitting up the tuning data using TimeSeriesSplit, an exhaustive search for the optimal parameter combinations is performed through the grid search, which ranks parameters based on a scoring function. This study uses the out-of-sample RMSE as the scoring function to penalize large errors to improve prediction accuracy. By exhaustively searching for parameter combinations, the method is more robust compared to a random search method which only tests a limited number of combinations, with the downside that it is more resource-intensive (Boehmke & Greenwell, 2019; De Prado, 2018). The tuning process begins by optimizing the model to minimize the RMSE at the first training iteration. Following this initial tuning, the model is re-tuned every five years using the expanded training data.

Table 2. Hyperparameter tuning for gradient boosting model to minimize RMSE across four periods (1999-2023). The third column indicates the parameter ranges included in the tuning process. The subsequent four columns specify the selected parameters for each period, where the training data for Period 1 covers 1999-2008, Period 2 covers 1999-2013, Period 3 covers 1999-2018 and Period 4 covers 1999-2023.

Parameter	Definition	Part of Tuning	Period 1	Period 2	Period 3	Period 4
n_estimators	Number of trees in the model	100 - 3000	500	200	100	200
learning_rate	Step size shrinkage to prevent overfitting	0.01 - 0.1	0.01	0.05	0.01	0.01
max_depth	Maximum depth of each tree	1 - 5	1	1	3	3
subsample	Fraction of samples used for training each tree	0.5 - 1	0.8	0.6	0.5	0.6
colsample_bytree	Fraction of features used per tree	0.5 - 1	0.6	0.6	0.6	0.6
lambda	L2 regularization term on weights	0 - 1	1	0.5	1	1
alpha	L1 regularization term on weights	0 - 1	0.5	0.5	0	0.5

Table 2 shows the range within which parameters are tuned and tested, which is decided based on running the model repeatedly and evaluating its performance. The selected parameters resulting from the tuning process are also shown for all four periods, with most parameters being tuned differently across periods.

4.2.4 Out-of-Sample Estimation

To evaluate the predictability of commodities, out-of-sample R^2 is used as an evaluation metric. While in-sample R^2 measures how well the model explains the variability of the data it was trained on (the training set), out-of-sample R^2 measures how well the model predicts new, unseen data (the validation set). It is calculated as follows:

$$R^2_{i,OOS} = 1 - \frac{\sum_{t=s}^{T}\left(y_{i,t} - \hat{y}_{i,t}\right)^2}{\sum_{t=s}^{T}\left(y_{i,t} - \bar{y}_{i,t}\right)^2}, \tag{11}$$

where $\hat{y}_{i,t}$ is the conditional forecast, $y_{i,t}$ is the predicted value of the stock index log return for the country i at time t and $\bar{y}_{i,t}$ is the unconditional forecast which is the historical means of log returns over the training period. If $R^2_{i,OOS} > 0$, the conditional forecast outperforms the unconditional one, indicating that the features have predictive powers.

4.2.5 Pooled Estimation

A pooled out of sample R^2 metric is also used, where the squared errors are aggregated across all indices, treating the predictions as a single dataset. This methodology follows the approach by Hjalmarsson (2010), who advocates pooling data across multiple entities, such as countries or assets, to enhance the robustness of predictive regressions. Unlike the traditional out-of-sample R^2 which trains separate models for each stock index, the pooled approach uses one model to capture shared patterns and correlations among the indices, potentially improving predictive performance and increasing robustness. Hjalmarsson provides both practical and economic justifications for pooling, which we adapt to our GBDT framework; if the predictive relationships are similar across indices, pooling the data can yield more precise estimates than individual models. Even when these relationships differ, the pooled model converges to an average predictive relationship, offering a useful summary of predictability. Economically, Hjalmarsson notes that similar predictability patterns are more likely in developed economies due to integrated financial markets, whereas emerging markets may exhibit unique return profiles. In our study, we hypothesize that commodity prices can influence global stock prices through shared macroeconomic effects, justifying the pooling of data across indices to leverage these correlations and enhance model robustness compared to estimating 39 separate models.

4.2.6 Feature importance

Feature importance for the model is estimated using SHAP values, which derive from concepts in game theory. SHAP values compare each feature’s contribution to the predicted value against the average prediction for the entire dataset and a SHAP value is assigned accordingly. This method is found to provide several advantages to traditional approaches, such as the XGBoost built-in gain feature importance method, especially in datasets with correlated features. The strong performance comes from accounting for feature dependencies. For example, a feature might have a high feature importance using the gain method because it is included in many splits, but its actual contribution to the prediction might be small if those splits do not significantly change the outcome. SHAP values, by contrast, provide a more accurate measure by calculating the contribution of each feature to each prediction with respect to dependencies and aggregating these contributions for global importance. A drawback of choosing Shapley values is a significant increase in computation time (Lundberg & Lee, 2016; Lundberg et al., 2019). Nevertheless, due to the correlation and interactions between commodities in the dataset, SHAP values are used for evaluating the feature importance of the model. Gain feature importance will also be measured and compared to the SHAP values for robustness purposes and to identify any differences that may occur due to feature dependencies.

4.3 Trading strategy

To apply the insights of commodities' predictive power on stock markets, we propose the following simple trading strategy and evaluate whether it can be used to generate superior returns for investors. In line with the MPT, the objective is to achieve competitive mean-variance performance, as measured by the Sharpe ratio.

4.3.1 Model Training

Following the same GBDT model framework as the initial model, a trading strategy is constructed in Python based on a model trained on the same 1999–2023 training data and simulates trades for 2024. For each of the 39 stock indices, a separate XGBoost regressor is trained using the 16 commodity features in our sample, shifted by one day to avoid look-ahead bias. A 2-day lag, 3-day lag, 5-day rolling mean, and a 5-day momentum feature are added for all commodities, increasing the total number of features to 80. The indices' log returns are set as the target variable. Hyperparameters are tuned via the grid search function using directional accuracy as the scoring function.

Pardo (2008) argues that the optimal trading strategy should not be based on returns or the Sharpe ratio alone. Instead, more variables need to be considered for a robust trading strategy to differences in market conditions and outliers. The key characteristics he suggests are relatively even distribution of trades, relatively even distribution of trading profit, relative balance between long and short profit, a large group of contiguous and profitable strategy parameters in the optimization, acceptable trading performance in a wide range of markets, acceptable risk, relatively stable winning and losing runs, and a large and statistically valid number of trades (Pardo, 2008). Pardo's insights are considered by evaluating the distribution and consistency of returns in conjunction with the Sharpe ratio and annualized net return. This is done through visual observation of the output of using different numbers of hyperparameters, daily positions, and minimum prediction R^2 levels. For example, a strategy that only includes higher confidence trades (through a higher minimum R^2 requirement) may result in increased annualized return and Sharpe ratio but still be neglected due to only a limited number of trades being undertaken, making the strategy less robust. The strategy's performance will be evaluated using the Sharpe ratio, which will be compared to benchmarks. A long-short strategy is proposed to remove systematic risk and emphasize the commodity-stock relationship. With the same pool of indices being used for the long/short positions and an equal number of long and short positions, the beta of the trading portfolio is expected to produce a beta around zero. For this reason, the strategy's performance will be benchmarked against peer market-neutral strategies rather than the market portfolio. The long/short strategy is also dollar-neutral and thus requires no investment other than transaction costs.

4.3.2 Trading Simulation

The trained GBDT model simulates a market-neutral trading strategy for 2024 to evaluate its performance from January 2 to December 31. On each trading day, the following steps are executed. First, a prediction is made based on the predict_next_day function in Python, which uses the trained model to forecast the log returns of all 39 stock indices based on the feature values. A signal is then generated based on the predicted log returns. Indices with models having an R^2 score below 0.015 or an absolute expected change below 0.05 percent are excluded to ensure prediction reliability and trade quality, a precaution against overfitting (de Prado, 2018). Based on the signals, the three indices with the highest predicted percentage changes are selected for long positions, and the three with the lowest predicted changes are selected for short positions, resulting in a balanced portfolio of six trades per day, provided at least six indices meet the mentioned criteria. For each position, 10 000 dollars is allocated, and a transaction cost of one dollar (or one basis point) is deducted per trade. A risk-free rate of 4.2 percent is used based on the average US 3-month treasury yield. The daily total return is the sum of the six position returns net of transaction costs. A performance evaluation is conducted through summary statistics and benchmarked against peer market-neutral trading strategies.

5. Results and Analysis

This section begins by presenting empirical findings from the predictions of the GBDT model. Results are evaluated using in-sample, out-of-sample, and pooled out-of-sample performance metrics, followed by an analysis of the findings. The model is then further analyzed through a feature importance analysis using the SHAP and gain feature importance methods. Lastly, the performance of the trading strategy is demonstrated and evaluated.

5.1 In-sample, out-of-sample and pooled out-of-sample predictability

Tables A3 and A4 show the in-sample, out-of-sample and pooled out-of-sample R^2 values of 12 individual commodities and 4 commodity indices on the stock markets of 39 countries. To spare the reader from excessive scrolling to the abstract section for the complete tables, Tables 3 and 4 in section 5 presents subsets of them, which only include the most relevant countries relating to the analysis. A breakpoint value of 1 percent is used to better categorize the data, where R^2 values above 1 percent are labeled strong predictors and values below are labeled weak predictors. The commodities with the strongest evidence of predictability will be further analyzed, and Table A7 shows the production as well as consumption of these commodities for selected countries and is used to interpret the results. Heterogeneous effects of individual commodities and commodity indices are highlighted in the heatmaps presented in Figures 2, 3, and 4 in section 5. These figures provide insight into the geographical distribution of R^2 values to study the differences in predictability between countries and markets.

5.1.1 Individual commodities

The results for individual commodities are shown in Table A3. As seen in Panels A, B, D, G, I, J, K, and L, cattle, coffee, corn, gold, natural gas, silver, sugar, and wheat show almost no predictive power for stock markets, with few or zero R^2 values exceeding 1 percent. This is largely consistent with previous research discussed in Section 2.1 and 2.2, where these commodities are typically seen as receivers of volatility shocks from stocks rather than transmitters. Gold is the most extensively researched out of them, and our results do not contradict previous findings of gold's role as a safe haven (e.g., Dai et al., 2022; Mensi et al., 2024). While Iyke and Ho (2021) find some predictive power for these commodities depending on the market state, period, and evaluation metric, our results suggest that other factors primarily govern stock markets. Jordan et al. (2016) found predictability for gold and silver, but as they sampled the Canadian stock market, which is excluded from our sample, our results neither support nor contradict theirs.

Table 3: A subset of the data from Table A3 for selected countries. The table shows In-Sample, Out-of-Sample and Pooled results for copper and cotton. The first column displays the sampled country, followed by the 3 different evaluation metrics for each commodity panel. The R^2 values are listed as percentages.

Country	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)
	Panel C. Copper			Panel D. Cotton		
Argentina	0.900	0.220	0.651	0.675	0.235	0.456
Australia	5.107	4.992	4.455	2.151	1.762	1.429
Belgium	1.421	1.051	0.937	1.123	0.952	1.000
Finland	1.424	1.497	1.338	1.090	1.292	0.997
Hong Kong	2.527	2.953	2.764	1.120	1.004	0.962
Netherlands	1.469	1.353	1.299	1.297	1.157	1.151
Norway	2.810	1.716	1.684	1.734	1.220	0.957
Japan	2.953	2.256	2.143	1.056	0.389	0.425
Singapore	2.928	3.190	2.558	0.871	0.624	0.491
South Korea	2.379	3.050	2.676	0.647	0.852	0.847
Turkey	0.888	0.186	0.236	0.744	0.450	0.511
United States	0.827	0.650	0.782	0.481	0.480	0.570

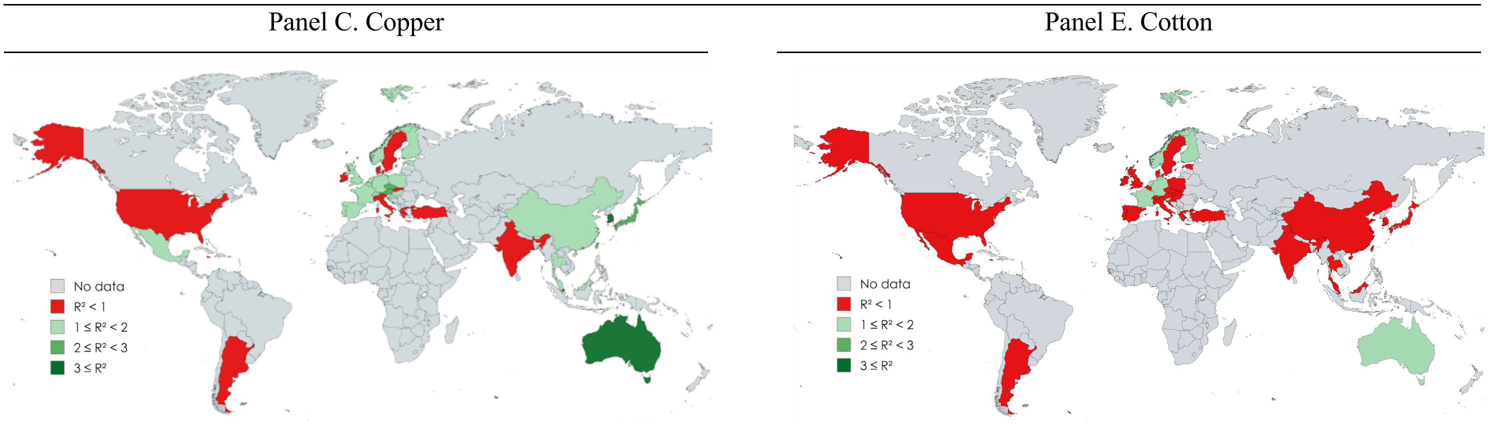


Figure 2: The heatmap illustrates the geographical distribution of out-of-sample R^2 values for copper and cotton on the stock markets of 39 countries. Panel C displays the distribution for copper, while Panel E shows the distribution for cotton. Grey colored countries are those not included in the sample. Red colored countries are countries with R^2 values < 1, while the green colored countries are those with R^2 values > 1 with darker shades representing higher values.

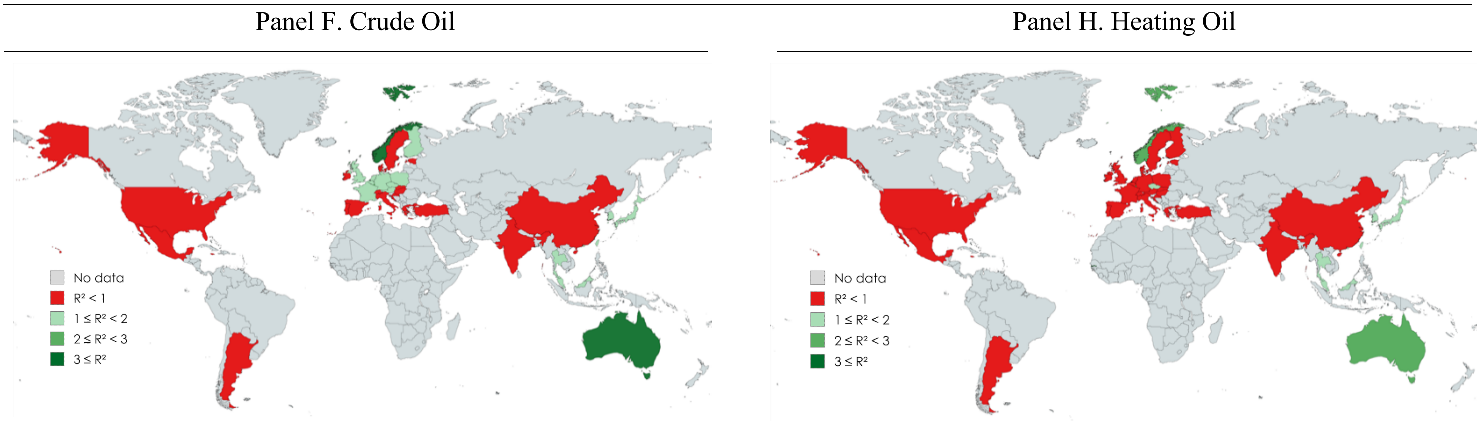
Copper shows strong in-sample R^2 values for 33 out of the 39 sampled countries, as 33 countries exceed the breakpoint R^2 level of 1 percent. 27 countries also report strong out-of-sample R^2 scores, with the highest ones being for Australia, Singapore, and South Korea. Pooled R^2 is strong for 21 countries, with Australia, Hong Kong, and South Korea having the highest values. The findings support strong explanatory power for copper on stock markets across all performance metrics. A subset of the data is shown in Table 3. Copper also has the highest singular R^2 scores for an individual commodity in our sample, with the R^2 for Australia being 4.992 percent out-of-sample and 4.455 percent in the pooled metric. Heterogeneous effects of copper price changes are observed at the country level, with R^2 values ranging from around zero to 4.992 percent out-of-sample and around zero to 4.455 percent pooled.

Figure 2 (Panel C) provides an overview of the geographical distribution of out-of-sample R^2 values for copper. The high explanatory power of copper for the Australian stock market (as illustrated by the darker shade of green) can be explained by the country's role as a major copper producer. The Australian Security Exchange, which is used as the representative index, includes mining giants such as BHP, Fortescue, and Rio Tinto, which further explains the result. Large copper consumers like South Korea and Japan also receive high R^2 values across all valuation metrics. Furthermore, highly globally integrated trade economies such as Hong Kong and Singapore show consistently high R^2 values, despite not being a major producer or consumer of the commodity. This result suggests spillover effects where changes in the copper price affect the global economy in a broad sense, supporting previous literature discussed in sections 2.1 and 2.2. Another pattern revealed in Figure 2 (Panel C) is that predictability is found for many European countries, supporting the hypothesis that copper affects highly integrated markets. It may also signify tightly integrated supply chains in Europe, where copper is a driver in industrial sectors such as construction and electronics. Some exceptions exist, however, with weak evidence of explanatory power being found for the US market. This suggests that despite being a major copper consumer and producer in nominal terms, the American market may be more driven by other sectors such as technology. Consistently low explanatory power was also found for countries such as Argentina and Turkey.

For cotton, strong in-sample R^2 values are found for 19 countries. The out-of-sample R^2 scores are strong for eight countries, with the highest scorers being Australia, Finland, and Norway. Pooled out-of-sample R^2 is strong for three countries: Australia, Belgium, and the Netherlands. Table 3 (Panel E) shows a subset of the data. While the findings suggest some explanatory power for cotton, it is mainly in-sample. The number of strong R^2 values decreases significantly in the out-of-sample and pooled metrics, suggesting a lack of robustness and overfitting on the in-sample evaluation. Figure 2 shows how some predictability is found for specific European countries and Australia. The highest predictability is once again found in Australia. Its position as a significant cotton producer helps explain its sensitivity to changes in cotton prices. Other countries with strong out-of-sample predictability include Belgium and the Netherlands, which are also robust in the pooled evaluation. The findings are similar to those of previous researchers discussed in section 2.1.

Table 4: A subset of the data from Table A3 for selected countries. The table shows In-Sample, Out-of-Sample and Pooled R^2 results for crude oil and cotton. The first column displays the sampled country, followed by the 3 different evaluation metrics for each commodity panel. The R^2 values are listed as percentages.

Country	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)
	Panel F. Crude Oil			Panel K. Heating Oil		
Australia	2.713	3.241	2.740	2.022	2.325	1.902
China	0.717	0.601	0.559	0.581	0.056	0.082
Hong Kong	1.426	1.718	1.991	1.223	1.258	1.423
Malaysia	1.447	1.703	1.730	1.342	1.302	1.155
Norway	3.793	3.526	3.906	2.811	2.153	2.517
Singapore	1.246	1.726	1.394	1.165	1.266	0.857
Spain	0.890	0.880	0.718	0.670	0.374	0.287
United States	0.579	0.177	0.341	0.545	0.156	0.239


Panel F. Crude Oil

Panel H. Heating Oil

Figure 3: The heatmap illustrates the geographical distribution of out-of-sample R^2 values for crude oil and heating oil on the stock markets of 39 countries. Panel F displays the distribution for crude oil, while Panel H shows the distribution for heating oil. Grey colored countries are those not included in the sample. Red colored countries are countries with R^2 values < 1, while the green colored countries are those with R^2 values > 1 with darker shades representing higher values.

Crude oil exhibits predictive power, with strong in-sample R^2 values for 20 countries. 18 countries also report strong out-of-sample R^2 values, with Australia, Norway, and Singapore showing the highest scores. The pooled performance is strong for 14 countries, with the highest R^2 values found for Australia, Norway, and Hong Kong. The strongest predictability is found for Norway which is a major oil producer. Australia also shows high R^2 values across all metrics, despite not being a major oil producer or consumer, unlike for copper and cotton. A potential explanation lies in the previous findings discussed in Section 2.2, which show that crude oil demonstrates a positive correlation with other commodities—especially copper—and that this correlation has increased in recent years. As with copper, high R^2 values for crude oil are observed for Hong Kong and Singapore. Singapore is a significant crude oil consumer, but they do not have any oil production, and neither does Hong Kong. A similar pattern is therefore observed for crude oil as for copper, where integrated trade economies are affected more strongly. While this may also be explained by strong correlation with copper, it can also suggest spillover effects for crude oil onto the global market. Another recurring pattern is the wide range of R^2 scores, with values ranging from around zero to 3.526 out-of-sample and around zero to 3.906 for the pooled metric for the sampled countries. The heterogeneous effects are highlighted in Figure 3 (Panel F) where the geographical distribution of the R^2 values for the sampled countries is shown. Predictability is found for stock markets in Europe, Asia, and Australia. Countries where no evidence of predictability include China, Spain and the US. The result aligns with previous research discussed in sections 2.1 and 2.2, where crude oil is typically labeled as a net transmitter of volatility to stock markets. It is also aligned with the findings of Fasanya et al. (2023) who found no predictability for the Chinese market.

For heating oil, strong in-sample R^2 scores are found for 12 countries. 10 countries also show strong out-of-sample predictability, with the largest R^2 values for Australia, Norway, and Malaysia. The pooled R^2 is strong for seven countries, with Norway, Australia, and Malaysia having the highest values. The countries that are affected the most by heating oil price changes are all markets that are shown to be impacted by price changes in crude oil, suggesting strong co-movement and integration between the two commodities. As heating oil is produced from crude oil, the large producing and consuming countries are the same for the two commodities. Figure 3 (Panel H) shows the geographical distribution of R^2 values for heating oil. A comparison between the heatmaps for crude oil (Panel F) and heating oil (Panel H) reveals how predictability for heating oil is mainly found in countries where the impact of crude oil is the highest.

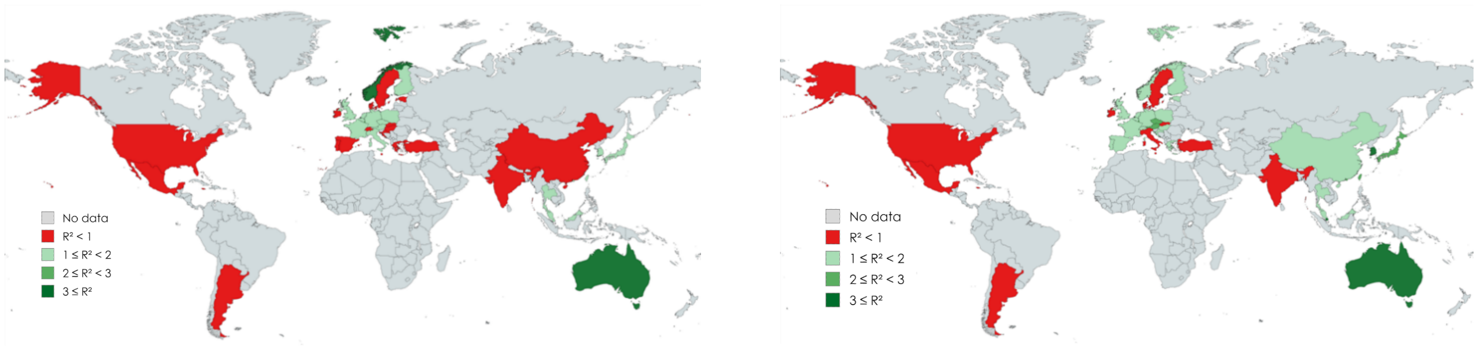
5.1.2 Commodity indices

The model output for commodity indices is shown in Table A4. Limited predictability is found for the indices for agriculture and precious metals, as shown in Panels A and D, respectively. Predictive power is found for 21 countries in-sample for the agriculture index. However, out-of-sample predictability is only strong for three countries: Australia, Taiwan, and Hong Kong. Malaysia is the only country with a strong pooled R^2. Poor translation from in-sample to out-of-sample performance suggests overfitting and a lack of robustness. This result is consistent with the results of Iyke and Ho (2021) who find no strong evidence of predictability for the agriculture index.

Precious metals have strong in-sample values for six countries. Slovakia is the only country with a strong out-of-sample and pooled out-of-sample R^2, with values of 1.289 and 1.554, respectively. These findings are in agreement with research by Jacobsen et al. (2019) who finds limited predictability for precious metals.

Table 5: A subset of the data from Table A4 for selected countries. The table shows In-Sample, Out-of-Sample and Pooled R^2 results for the commodity indices energy and industrial metals respectively. The first column displays the sampled country, followed by the 3 different evaluation metrics for each commodity index panel. The R^2 values are listed as percentages.

Country	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)
	Panel B. Energy			Panel C. Industrial Metals		
Australia	2.846	3.491	2.987	5.398	5.289	4.549
Hong Kong	1.469	1.822	2.034	2.600	2.973	2.914
Jamaica	0.348	-0.058	-0.160	0.284	0.027	-0.008
Japan	1.413	1.608	1.485	2.960	2.216	2.019
Norway	3.460	3.546	4.060	2.988	1.684	1.785
Singapore	1.354	1.736	1.302	3.173	3.363	2.901
South Korea	1.245	1.725	1.700	2.357	3.341	2.875
Turkey	0.940	-0.061	-0.089	0.916	0.031	0.122
United States	0.606	0.248	0.371	0.669	0.390	0.678



Panel B. Energy

Panel C. Industrial Metals

Figure 4: The heatmap illustrates the geographical distribution of out-of-sample R^2 values for the energy and industrial metals indices on the stock markets of 39 countries. Panel C displays the distribution for energy, while Panel E shows the distribution for industrial metals. Grey colored countries are those not included in the sample. Red colored countries are countries with R^2 values < 1, while the green colored countries are those with R^2 values > 1 with darker shades representing higher values.

The energy index shows signs of predictability, with strong in-sample R^2 scores for 18 countries. Out-of-sample scores are strong for 19 countries, with Norway, Australia, and Hong Kong showing the largest values. The pooled R^2 metric is strong for 16 countries, with the same countries displaying the largest values. Table 5 shows a subset of the results. Similar results are found for energy as for crude oil, with the same countries showing the highest predictability. Figure 4 (Panel B) shows a heatmap of the geographical distribution of out-of-sample R^2 values. Its strong resemblance to Figure 3 (Panel F) suggests that crude oil is a major driver of the energy index, with a similar distribution of heterogeneous effects on the countries in the sample. The findings of predictive power for the energy index align with the study by Iyke and Ho (2021). It is also consistent with the study by Jacobsen et al. (2019) who found no predictability for the energy index, as they only studied the US market. The conflicting results highlights the importance of studying the impact of commodities at the country-level, as their predictability on stock markets varies significantly depending on the geographical location of that market.

For industrial metals, R^2 scores are strong for 33 countries in-sample and 27 out-of-sample, with the largest values being found for Australia, Singapore, and South Korea. The pooled R^2 is strong for 24 countries, with Australia, Singapore, and South Korea again showing the highest predictability. The industrial metals index shows the strongest evidence of predictability out of all the sampled features, with the largest number of strong R^2 values as well as the highest individual

R^2 values. Countries with the highest R^2 values for copper are also the ones with the highest R^2 scores for the industrial metals index, with Australia, Hong Kong, Japan, Singapore, and South Korea showing the strongest evidence of predictability. Australia displays the highest R^2 values across all evaluation metrics, with values of 5.289 and 4.549 percent for the out-of-sample and pooled metric, respectively. The large weighting of the Australian Securities Exchange in mining again explains the result, and the strong resemblance to the results for copper suggests that the commodity is a major driver of the industrial metals index. This co-movement is further illustrated in Figure 4 (Panel C), which shows that the geographical distribution of R^2 values is almost identical to Figure 3 (Panel C). Strong predictability is found, just as for copper, in most of the sampled countries in Europe and Asia, as well as Australia. No evidence of predictability is found for Mexico, Turkey and the US. The findings of strong predictability for industrial metals are consistent with Black et al. (2014) and Jacobsen et al. (2019).

5.2 Feature importance

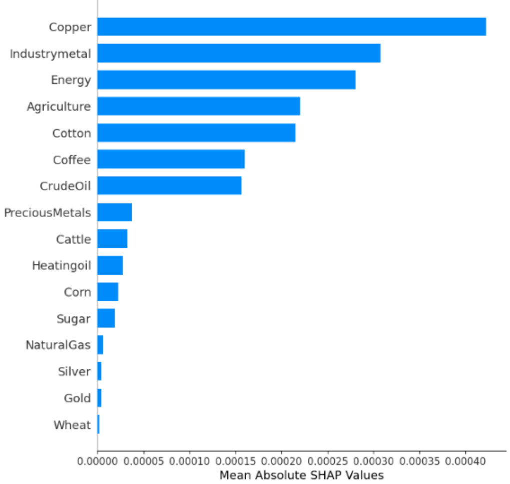


Figure 5: Bar chart measuring Mean Absolute SHAP Feature Importance. It demonstrates the average absolute impact of each feature on the predictions of the GBDT model and shows their relative importance. The x-axis represents the mean impact of log commodity returns on the prediction of log stock returns, and a value of 0.00043 for copper indicates an average impact of ~0.043 percent on the predictions for all indices.

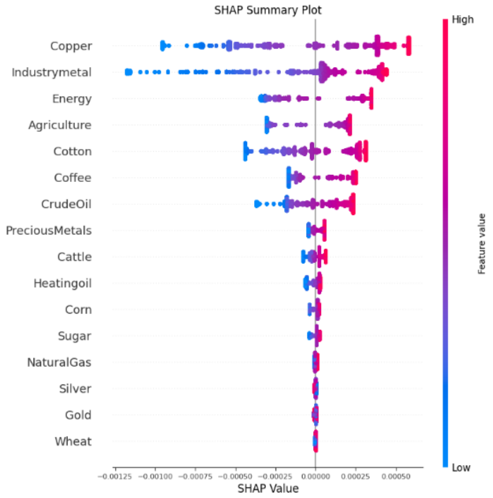


Figure 6: SHAP Summary Plot. The plot shows feature importance and the directional impact of the features, where each dot represents a single prediction. The x-axis represents the magnitude and direction of each feature's contribution to the predictions. High commodity values are illustrated by red dots, and red dots paired with positive SHAP values suggests a positive relationship. One dot with a SHAP value of 0.0005 for copper indicates a positive impact of ~0.05 percent in predicted log return for the index for that prediction.

Figures 5 and 6 show SHAP values for the sampled features. Figure 5 presents the mean absolute SHAP values, where Copper shows the highest mean absolute SHAP value of around 0.00043, meaning that copper prices contribute 0.043 percent to the variation in predicted stock returns on average for all the indices. The small magnitude of the SHAP values has two explanations. First, it estimates the impact on daily stock return movements, where small percentage movements is to be expected. Second, the impact is diluted by the fact that several indices show no evidence of predictability, which lowers the average impact. After copper, the indices of industrial metals, energy, and agriculture display the highest feature importance in predictions, followed by cotton, coffee, and crude oil. The other commodities do not seem to have any significant impact on the predictions, which is consistent with the R^2 values given in Tables A3 and A4. The only exception is coffee, which appears to have some importance in predictions through its interaction with other features, despite not showing any individual predictive power.

Figure 6 presents the SHAP summary plot across all indices, illustrating both feature importance and the directional impact of the features. Each dot represents a single prediction, where the x-axis shows the impact of a feature on the model's predicted return, and the color gradient reflects the feature value which is the log return of each commodity. Positive SHAP values indicate that a feature is associated with an increase in the index predicted return, while negative values are associated with a decrease. For copper, high values, as measured by red dots, are associated with positive SHAP values, indicating that copper prices tend to push stock predictions upward. Conversely, low copper values, as measured by blue dots are associated with negative SHAP values. A similar positive relationship is found for all the significantly large features.

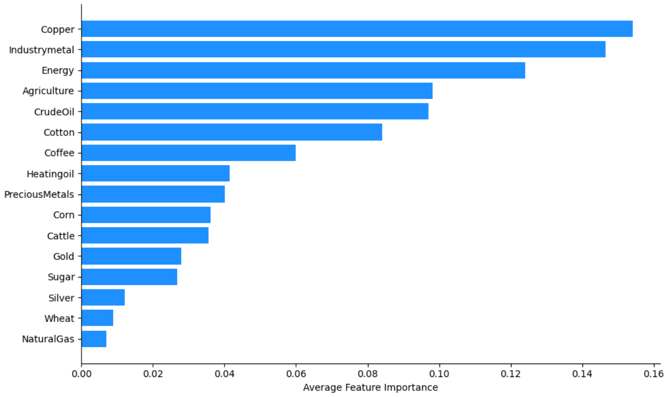


Figure 7: Average Feature importance of the features using XGboost's gain method. The bar chart shows the average improvement in the model's loss function. The x-axis measures the relative contribution, and the values sum to 1. A higher value indicates a higher feature importance.

Figure 7 shows the feature importance using the gain method, which shows the relative contribution of the sampled commodities to reducing the loss function of the XGBoost model. The x-axis is therefore only interpretable in relative terms unlike in the SHAP figures, and the values sum to 1. The similar relative feature importance ranking of commodities as in the SHAP analysis in Figures 5 and 6 provide support for its robustness. Small differences include lower importance for coffee and higher importance for gold compared to SHAP outputs. These discrepancies can be explained by the simpler computation of the gain method which doesn't account for feature dependencies, unlike SHAP which captures the interactions between features. Notably, despite not showing any predictive power in Panel B of Table A3, coffee shows some importance in both SHAP and gain method analyses.

5.3 Trading strategy

Subsets of the data output for the trading strategy is depicted in Tables A5 and A6, showing the first couple of days/weeks of trading. The trading period is ongoing from January 2 to December 31, and trades only occur on trading days. Table A5 depicts how a trading day involves three long positions, and three short positions based on a signal, which in turn depends on the predicted percentage return of the index. It also shows that a diverse set of indices is utilized for the positions. The confidence of each prediction is measured by R^2 in the rightmost column. Table A6 shows what the total daily net return looks like for the first trading month, with some days reporting losses.

Table 6: Summary statistics of the trading strategy. The table reports its performance through key performance metrics.

Metric	Value
Number of Trades	1512
Total Return	$6548
Average Return per Trade	$4.30
Sharpe Ratio (RFR = 4.2%)	0.83
Annualized Return	10.91%
Annualized Standard Deviation	8.08%
Beta	0.01

Table 7: Sharpe Ratio Sensitivity Analysis Table. It shows how the Sharpe Ratio varies with different risk-free rates and trading costs, where the trading cost is the cost of each buy/sell transaction in dollars.

		Trading Cost ($)				
		$0	$0.5	$1	$1.5	$2
Risk-Free Rate	3.70%	1.16	1.02	0.88	0.75	0.61
	3.95%	1.13	0.99	0.86	0.72	0.58
	4.20%	1.11	0.96	**0.83**	0.69	0.56
	4.45%	1.08	0.94	0.8	0.67	0.53
	4.70%	1.05	0.91	0.77	0.64	0.5

A summary of the performance for the entire trading year is given in Table 6. The output shows that the strategy successfully identifies trading opportunities on each trading day, as six positions on each of the 252 trading days sum up to 1512. Total return for the trading year is 6548 dollars making the average return per trade 4.3 dollars. An annualized return of 10.91 percent and a standard deviation of 8.08 percent are achieved, resulting in a Sharpe ratio of 0.83. Our reported Sharpe ratio falls within the range of 0.59 to 1.45 achieved by peer market-neutral studies using alternative predictors (e.g., Asness et al., 2013; Blitz et al., 2019; Gatev et al., 2006; Lin et al., 2021).

Table 7 presents the sensitivity of the performance of the trading strategy, as measured by the Sharpe ratio, to variations in transaction costs and risk-free rates. An increase in either transaction costs or the risk-free rate results in a decline in the Sharpe ratio. In particular, trading costs have a large impact on profitability and thus on the competitiveness of the trading strategy.

6. Conclusion

We analyze the predictability of stock returns using commodity futures and index prices across a dataset covering 39 countries from 1999 to 2024, by employing the XGBoost implementation of the GBDT approach. Our results build on existing literature by applying machine learning techniques and uncovering heterogeneous effects across countries.

Regarding the first research question, our findings support the ability of GBDT to predict stock returns using commodities. Strong evidence of predictability is found, most notably for copper, which shows R^2 values exceeding 1 percent for 69 percent and 54 percent of countries for the out-of-sample and pooled metric, respectively. Predictive abilities are also observed for cotton, crude oil, and heating oil. For the four commodity indices, industrial metals show the best predictive abilities, with R^2 values exceeding 1 percent for 69 percent and 62 percent of countries for the out-of-sample and pooled metrics, respectively. Our results also suggest strong predictability for energy, but weak predictive performance for agriculture and precious metals.

For the second research question, we provide evidence of heterogeneous effects, as we find a large variation in the exposures of countries to changes in commodity prices. For copper, R^2 values range from around zero for Argentina, Jamaica, and Turkey to 4.992 out-of-sample and 4.549 pooled for Australia. A similar heterogeneous impact appears for crude oil, with R^2 values around zero for Estonia, Malta, and Slovakia to 3.526 out-of-sample and 3.906 pooled for Norway. The large variation is explained by the difference in commodity exposure, with Australia heavily tied to mining industries and Norway to oil. The stock markets of globally integrated trade hubs like Singapore and Hong Kong also exhibit high sensitivity to the commodity price changes, despite limited direct exposure, suggesting spillover effects of commodities onto the global economy. For several countries such as Argentina, Turkey and US, no predictive power is found for any commodity, suggesting that their economies are driven by other factors than commodity prices. The out-of-sample and pooled results indicate that copper and crude oil's predictive ability is comparable to established predictors like interest rates, dividend yields, and US lagged returns for some countries and markets. The predictability appears to be highest for countries that are major producers or consumers of the commodity or globally integrated trade economies, although a more in-depth analysis into the root causes of the commodity exposures of each country is beyond the scope of this study and will be left for future researchers.

The third research question is also explored, and a competitive trading strategy is implemented based on predicting stock returns using commodities. The strategy is based on the same GBDT framework but tuned separately to fit the proposed long-short market-neutral setting. Achieving a Sharpe ratio of 0.83, its performance is comparable to benchmark peer market-neutral strategies using other predictors. Its competitiveness is highly sensitive to transaction costs which must be evaluated by the investor before implementation.

6.1 Limitations and Future Extensions

Finally, we address the limitations of our study and areas for future research. There are multiple directions for further work to extend our research on the predictability of commodities on stock returns, the most notable of which is expanding the sample size to include more countries. This will become easier as stock market data availability improves for developing countries. In the same way, data availability of individual commodity futures is set to improve, and future researchers may look to explore more of them. Additionally, while previous papers have used econometric models to find shifting impacts of commodities on stock markets depending on factors such as market expansions/recessions, breakpoints, and asymmetries, these are not specifically addressed in our machine learning setting. Instead, our model is trained in periods that include these scenarios to try to learn from them for future predictions. The practitioner and reader of our machine learning method remain uninformed of these conditional and state-switching properties despite them being accounted for in the model's predictions. Another area for future research to explore is implementing more frequent hyperparameter tuning. This increases computational power requirements but can improve the predictive accuracy to a degree.

The trading strategy is purposefully kept simple for clarity and the prevention of data mining, and its suitability for practical implementation requires a more detailed breakdown of transaction costs. It is also subsidiary to the main purpose of this study, which is to evaluate predictability. Future researchers may build on our work by exploring more advanced trading strategies using our findings on commodity-stock predictability.

References

Aizenman, J., Lindahl, R., Stenvall, D., & Uddin, G. S. (2024). Geopolitical shocks and commodity market dynamics: New evidence from the Russia-Ukraine conflict. *European Journal of Political Economy, 85,* 102574.

Ali, S., Bouri, E., Czudaj, R. L., & Shahzad, S. J. H. (2020). Revisiting the valuable roles of commodities for international stock markets. *Resources Policy*, *66,* 101603.

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

Arfaoui, M., & Ben Rejeb, A. (2017). Oil, gold, US dollar and stock market interdependencies: a global analytical insight. *European Journal of Management and Business Economics, 26*(3), 278-293.

Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum everywhere. *The journal of finance, 68*(3), 929-985.

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

Biswas, P., Jain, P., & Maitra, D. (2024). Are shocks in the stock markets driven by commodity markets? Evidence from Russia-Ukraine war. *Journal of Commodity Markets, 34*, 100387.

Black, A. J., Klinkowska, O., McMillan, D. G., & McMillan, F. J. (2014). Forecasting stock returns: do commodity prices help?. *Journal of Forecasting, 33*(8), 627-639.

Blitz, D., Van Vliet, P., & Baltussen, G. (2019). The Volatility Effect Revisited. *The Journal of Portfolio Management, 46*(2).

Boehmke, B., & Greenwell, B. M. (2019). *Hands-On Machine Learning with R*. Chapman and Hall/CRC.

Brown, S. (2021). Machine learning, explained. MIT Sloan. *Management Sloan School.* (Retrieved February 20, 2025, from https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained).

Cakici, N., Fieberg, C., Metko, D., & Zaremba, A. (2023). Machine learning goes global: Cross-sectional return predictability in international stock markets. *Journal of Economic Dynamics and Control, 155*, 104725.

Campbell, J. Y., & 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.

Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of financial economics, 18*(2), 373-399.

Campbell, J. Y., & Shiller, R. J. (1988). Stock prices, earnings, and expected dividends. *the Journal of Finance, 43*(3), 661-676.

Carmona, P., Dwekat, A., & Mardawi, Z. (2022). No more black boxes! Explaining the predictions of a machine learning XGBoost classifier algorithm in business failure. *Research in International Business and Finance*, *61*, 101649.

Chen, J. (2022). *Doctor Copper: Definition, theory, use as an indicator*. Investopedia. (Retrieved February 20, 2025, from https://www.investopedia.com/terms/d/doctor-copper.asp).

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

Dai, Z., Zhu, H., & Zhang, X. (2022). Dynamic spillover effects and portfolio strategies between crude oil, gold and Chinese stock markets related to new energy vehicle. *Energy Economics, 109,* 105959.

De Prado, M. L. (2018). *Advances in financial machine learning*. John Wiley & Sons.

Fama, E. F. (1965). The behavior of stock-market prices. *The journal of Business, 38*(1), 34-105.

Fama, E. F. (1970). Efficient Capital Markets. *Journal of finance, 25*(2), 383-417.

Fama, E. F., & French, K. R. (1988). Dividend yields and expected stock returns. *Journal of financial economics, 22*(1), 3-25.

Fama, E. F., & French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of financial economics*, *25*(1), 23-49.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, *81*(3), 607-636.

Fasanya, I. O., Adekoya, O., & Sonola, R. (2023). Forecasting stock prices with commodity prices: New evidence from Feasible Quasi Generalized Least Squares (FQGLS) with non-linearities. *Economic Systems, 47*(2), 101043.

Frazzini, A., Israel, R., & Moskowitz, T. J. (2012). Trading costs of asset pricing anomalies. *Fama-Miller Working Paper, Chicago Booth Research Paper*, (14-05).

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of* *statistics,* 1189-1232.

Garcia-Jorcano, L., & Sanchis-Marco, L. (2022). Spillover effects between commodity and stock markets: A SDSES approach. *Resources Policy*, *79*, 102926.

García, M. V., & Aznarte, J. L. (2020). Shapley additive explanations for NO2 forecasting. *Ecological Informatics, 56*, 101039.

Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. *The review of financial studies*, *19*(3), 797-827.

Geng, J. B., Chen, F. R., Ji, Q., & Liu, B. Y. (2021). Network connectedness between natural gas markets. *Energy Economics*, *95*, 105001.

Gong, X., Liu, Y., & Wang, X. (2021). Dynamic volatility spillovers across oil and natural gas futures markets based on a time-varying spillover method. *International Review of Financial Analysis, 76,* 101790.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223-2273.

Hjalmarsson, E. (2010). Predicting global stock returns. *Journal of Financial and Quantitative Analysis, 45*(1), 49-80.

Investing.com. (n.d.). *Commodities*. (Retrieved February 20, 2025, from https://www.investing.com/commodities).

Iyke, B. N., & Ho., S. Y. (2021). Stock return predictability over four centuries: The role of commodity returns. *Finance Research Letters, 40,* 101711.

Jacks, D. S., & Stuermer, M. (2020). What drives commodity price booms and busts?. *Energy Economics, 85,* 104035.

Jacobsen, B., Marshall, B. R., & Visaltanachoti, N. (2019). Stock market predictability and industrial metal returns. *Management Science,.65*(7), 3026-3042.

Jordan, S. J., Vivian, A., & Wohar, M. E. (2016). Can commodity returns forecast Canadian sector stock returns?. *International Review of Economics & Finance*, *41*, 172-188.

Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics*, *74*(2), 209-235.

Liao, J., Zhu, X., & Chen, J. (2021). Dynamic spillovers across oil, gold and stock markets in the presence of major public health emergencies. *International Review of Financial Analysis, 77, 101822.*

Lin, Y., Liu, S., Yang, H., Wu, H., & Jiang, B. (2021). Improving stock trading decisions based on pattern recognition using machine learning technology. *PloS one, 16*(8), e0255558.

Lundberg, S., & Lee, S. I. (2016) An unexpected unity among methods for interpreting model predictions. *arXiv preprint arXiv:1611.07478*.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., … & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence, 2*(1), 56-67.

Machine learning plus. (2021). *An introduction to gradient boosting decision trees.* (Retrieved February 23, 2025, from https://www.machinelearningplus.com/machine-learning/an-introduction-to-gradient-boosting-decision-trees/)

Markowitz, H. (1952). Portfolio Selection. The *Journal of Finance*, 7(1), 77-91.

Mensi, W., Rehman, M. U., Maitra, D., Al-Yahaee, K. H., & Vo, X. V. (2021). Oil, natural gas and BRICS stock markets: Evidence of systemic risks and co-movements in the time-frequency domain. *Resources Policy, 72,* 102062.

Naeem, M. A., Hamouda, F., & Karim, S. (2024). Tail risk spillover effects in commodity markets: A comparative study of crisis periods. Journal of Commodity Markets, 33, 100370.

NVIDIA. (2025). *XGBoost.* (Retrieved February 25, 2025, from https://www.nvidia.com/en-us/glossary/xgboost/).

OEC. (n.d.). *Russia.* (Retrieved February 7, 2025, from https://oec.world/en/profile/country/rus)

OEC. (n.d.). *Ukraine.* (Retrieved February 7, 2025, from https://oec.world/en/profile/country/ukr)

OECD. 2022. *The implications for OECD regions of the war in Ukraine.* (Retrieved February 20, 2025, from https://www.oecd.org/content/dam/oecd/en/publications/reports/2022/07/the-implications-for-oecd-regions-of-the-war-in-ukraine_09c96fae/8e0fcb83-en.pdf)

Pardo, R. (2011). *The evaluation and optimization of trading strategies*. John Wiley & Sons.

Rapach, D. E., Strauss, J. K., & Zhou., G. (2013). International stock return predictability: what is the role of the United States?. *The Journal of Finance*, *68*(4), 1633-1662.

S&P Global. (n.d.). *S&P Capital IQ Pro.* https://www.capitaliq.spglobal.com/

Salisu, A. A., Isah, K. O., & Raheem, I. D. (2019). Testing the predictability of commodity prices in stock returns of G7 countries: Evidence from a new approach. *Resources Policy, 64*, 101520.

Standard & Poor's. (n.d*.). Compustat Global - Daily Index prices* [Close] [1999-2024]. Wharton

Research Data Services. (Retrieved February 21, 2025, from https://wrds-www.wharton.upenn.edu/pages/get-data/compustat-capital-iq-standard-poors/compustat/global-daily/index-prices-daily/).

Shahzad, S. J. H., Bouri, E., Roubaud, D., & Kristoufek, L. (2020). Safe haven, hedge and diversification for G7 stock markets: Gold versus bitcoin. *Economic Modelling, 87*, 212-224.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance* , *19*(3), 425-442.

Sharpe, W. F. (1966). Mutual Fund Performance. *The Journal of business*, *39*(1), 119-138.

Solnik, B. (1993). The performance of international asset allocation strategies using conditioning information. *Journal of Empirical Finance*, *1*(1), 33-55.

Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting, 23,* 405–430.

Tang, K., & Wei, X. (2012). Index investment and the financialization of commodities. *Financial Analyst Journal, 68*(6), 54-74.

Thakkar, A., & Chaudhari, K. (2021). Fusion in stock market prediction: A decade survey on the necessity, recent developments, and potential future directions. *Information Fusion, 65,* 95-107.

U.S. Department of Agriculture, Foreign Agricultural Service. (n.d.). *Production: Cotton.* (Retrieved May 9, 2025, fromhttps://www.fas.usda.gov/data/production/commodity/2631000)

U.S. Energy Information Administration. (n.d.). *What drives crude oil prices?*. (Retrieved May 2, 2025, from https://www.eia.gov/finance/markets/crudeoil/financial_markets.php

U.S. Geological Survey. (2025). *Mineral commodity summaries 2025.* https://pubs.usgs.gov/periodicals/mcs2025/mcs2025.pdf

Wang, Y., Pan, Z., Liu, L., & Wu, C. (2019). Oil price increases and the predictability of equity premium. *Journal of Banking & Finance, 102*, 43-58.

World Population Review. (n.d.). *Oil production by country*. (Retrieved May 9, 2025, from https://worldpopulationreview.com/country-rankings/oil-production-by-country)

Ying, X. (2019). An overview of overfitting and its solutions. *Journal of Physics: Conference Series (Vol. 1168*, p. 022022). IOP Publishing

Zhou, F., Zhang, Q., Sornette, D., & Jiang, L. (2019). Cascading logistic regression onto gradient boosted decision trees for forecasting and trading stock indices. *Applied Soft Computing, 84,* 105747.

Appendix

Table A1: Stock Market Indices and corresponding Stock Exchange MIC Codes by Country. The first column indicates the country, and the second specific the stock market index associated with each country, the third and the fourth column represents the name of the stock exchange where the index is traded and the market identifier code (MIC) respectively.

Country	Index	Stock Exchange	MIC
Argentina	S&P Merval Index	Buenos Aires Stock Exchange	XMERV
Australia	S&P/ASX 300	Australian Securities Exchange	XASX
Austria	Austrain Traded Index	Vienna Stock Exchange	XWBO
Belgium	BEL 20 GR	Euronext Brussels	XBRU
China	SSE Composite Index	Shanghai Stock Exchange	XSHG
Croatia	CROBEX	Zagreb Stock Exchange	XZAG
Czech Republic	PX Index	Prague Stock Exchange	XPRA
Denmark	OMX Copenhagen 20	Nasdaq Copenhagen	XCSE
Estonia	Estonia Stock Market Index	Nasdaq Tallinn	XTAL
Finland	Helsinki General Index (HEX)	Nasdaq Helsinki	XHEL
France	CAC 40	Euronext Paris	XPAR
Germany	Deutscher Aktienindex (DAX)	Frankfurt Stock Exchange	XETR
Greece	Athens Stock Exchange General Index	Athens Stock Exchange	XATH
Hong Kong	Hang Seng Index	Hong Kong Stock Exchange	XHKG
Hungary	Budapest Stock Index	Budapest Stock Exchange	XBUD
India	BSE SENSEX	Bombay Stock Exchange	XBOM
Ireland	The Ireland Stock Market (ISEQ)	Euronext Dublin	XISE
Italy	FTSE MIB INDEX	Borsa Italiana	XMIL
Jamaica	JSE All Jamaican Composite Index	Jamaica Stock Exchange	XJAM
Japan	Topix Index	Tokyo Stock Exchange	XTKS
Luxembourg	Luxembourg Stock Market	Luxembourg Stock Exchange	XLUX
Malaysia	KLSE Composite Index	Bursa Malaysia	XKLS
Malta	Malta Stock Exchange	Malta Stock Exchange	XMAL
Mexico	IPC Mexico	Mexican Stock Exchange	XMEX
Netherlands	AEX-Index	Euronext Amsterdam	XAMS
Norway	Oslo Bors All-Share Index	Oslo Stock Exchange	XOSL
Poland	WIG Index	Warsaw Stock Exchange	XWAR
Portugal	PSI-20	Euronext Lisbon	XLIS
Singapore	Straits Times Index	Singapore Exchange	XSES
Slovakia	Slovakian Stock Market (SAX)	Bratislava Stock Exchange	XBRA
South Korea	Korea Stock Exchange Composite	Korea Exchange	XKRX
Spain	IBEX 35 Index	Madrid Stock Exchange	XMAD
Sweden	OMX Affarsvarlden Gernalindex	Nasdaq Stockholm	XSTO
Switzerland	Swiss Performance Index	SIX Swiss Exchange	XSWX
Taiwan	TAIEX	Taiwan Stock Exchange	XTAI
Thailand	SET Index	Stock Exchange of Thailand	XBKK
Turkey	BIST 100	Borsa Istanbul	XIST
United Kingdom	S&P United Kingdom	London Stock Exchange	XLON
United States	NYSE Composite	New York Stock Exchange	XNYS

Table A2: Commodity Futures and Commodity Indices. Panel A lists the commodity futures, with the first column indicating the commodity, the second column specifying the future contract, and the third column providing the ticker code for the futures. Panel B presents the commodity indices, where the first column denotes the category, the second column identifies the index, and the third lists the ticker.

Panel A. Commodity Futures		
Commodity	Future Contract	Ticker
Cattle	Live Cattle	LCc1
Coffee	US Coffee C	KCN5
Copper	Copper	HGN5
Corn	US Corn	ZCN5
Cotton	US Cotton No. 2	CTN5
Crude Oil	Crude Oil WTI	OIL
Gold	Gold	GCM5
Heating Oil	Heating Oil	NYFM5
Natural Gas	Natural Gas	NGM5
Silver	Silver	SIN5
Sugar	US Sugar No. 11	SBN5
Wheat	US Wheat	ZWN5
Panel B. Commodity Indices		
Category	Index	Ticker
Agriculture	S&P GSCI Agriculture	SPGSAG
Energy	S&P GSCI Energy	SPGSEN
Industry Metals	S&P GSCI Industrial Metals	SPGSIN
Precious Metals	S&P GSCI Precious Metals Index	SPGSPM

Table A3: In-Sample, and Out-of-Sample and Pooled R^2 results for three individual commodities The first column indicates the country, followed by 3 sets of columns for each commodity panel. For each panel, the table reports the in-sample R^2, out-of-sample R^2 and pooled out-of-sample R^2. The reported R^2 values are in percentages.

Country	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)
	Panel A. Cattle			Panel B. Coffee			Panel C. Copper		
Argentina	0.598	0.046	0.125	0.765	0.042	0.089	0.900	0.220	0.651
Australia	0.940	0.495	0.211	1.195	0.915	0.530	5.107	4.992	4.455
Austria	0.493	0.162	0.083	0.896	0.415	0.023	3.162	2.199	2.003
Belgium	0.419	0.113	0.074	0.766	0.504	0.194	1.421	1.051	0.937
China	0.372	-0.162	-0.243	0.662	0.350	0.306	1.507	1.231	1.336
Croatia	0.505	-0.178	-0.254	0.525	0.357	0.047	1.596	1.293	0.911
Czech Republic	0.813	0.165	-0.357	0.662	0.328	-0.022	2.995	2.037	1.673
Denmark	0.341	0.060	0.085	0.650	0.167	0.079	1.560	0.988	0.691
Estonia	0.415	-0.031	-0.280	0.803	0.587	0.248	2.223	1.774	0.951
Finland	0.564	0.085	0.113	0.538	0.222	-0.078	1.424	1.497	1.338
France	0.486	0.155	0.217	0.679	0.422	0.131	1.524	1.237	1.110
Germany	0.465	0.096	0.115	0.472	0.256	0.025	1.523	1.370	1.230
Greece	0.519	0.233	0.153	0.646	0.225	-0.072	1.532	0.898	0.956
Hong Kong	0.642	0.266	0.328	0.853	0.570	0.431	2.527	2.953	2.764
Hungary	0.472	-0.003	-0.012	0.541	0.123	-0.127	2.079	1.126	0.781
India	0.417	0.036	0.011	0.639	0.027	-0.309	0.968	0.833	0.837
Ireland	0.419	0.135	0.048	0.495	0.210	0.070	1.125	0.988	0.866
Italy	0.636	0.225	0.127	0.688	0.308	0.115	1.452	0.856	0.906
Jamaica	0.240	0.123	0.058	0.175	-0.073	-0.067	0.202	-0.004	0.008
Japan	0.508	0.041	-0.014	0.841	0.312	0.188	2.953	2.256	2.143
Luxembourg	0.473	0.155	0.112	1.076	0.618	0.318	2.206	1.793	1.635
Malaysia	0.410	0.114	0.004	1.082	0.975	0.943	1.684	1.743	2.095
Malta	0.339	-0.140	-0.220	0.627	0.269	0.210	1.048	0.562	0.377
Mexico	0.340	-0.097	-0.194	0.591	0.416	0.248	1.024	1.119	1.380
Netherlands	0.553	0.205	0.217	0.605	0.442	0.221	1.469	1.353	1.299
Norway	0.445	0.063	-0.024	0.904	0.550	0.147	2.810	1.716	1.684
Poland	0.537	-0.007	0.016	0.604	0.255	0.212	2.797	1.902	1.835
Portugal	0.472	0.119	0.081	0.661	0.332	0.178	1.590	1.185	1.076
Singapore	0.585	0.329	0.160	0.614	0.403	0.169	2.928	3.190	2.558
Slovakia	0.244	-0.115	-0.306	0.679	0.247	0.024	1.051	0.882	0.764
South Korea	0.596	-0.131	-0.206	0.972	0.809	0.132	2.379	3.050	2.676
Spain	0.516	0.160	0.148	0.646	0.362	0.127	1.547	1.137	0.939
Sweden	0.430	0.027	0.029	0.361	-0.039	-0.229	0.868	0.683	0.607
Switzerland	0.354	0.023	-0.017	0.598	0.326	0.018	1.203	1.027	0.760
Taiwan	0.716	0.194	0.253	0.856	0.840	0.443	1.611	2.751	2.268
Thailand	0.946	-0.060	-0.080	0.986	0.712	0.500	1.398	1.296	1.131
Turkey	0.437	-0.114	-0.172	1.073	0.227	0.292	0.888	0.186	0.236
United Kingdom	0.454	0.132	0.107	0.607	0.355	0.145	1.566	1.377	1.226
United States	0.318	-0.004	-0.018	0.392	0.121	0.093	0.827	0.650	0.782

Table A3: In-Sample R^2, and Out-of-Sample R^2 and Pooled out-of-sample R^2 results for three individual commodities (continued).

Country	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)
	Panel D. Corn			Panel E. Cotton			Panel F. Crude Oil		
Argentina	0.612	0.071	0.135	0.675	0.235	0.456	0.917	0.305	0.576
Australia	1.646	0.750	0.122	2.151	1.762	1.429	2.713	3.241	2.740
Austria	1.283	0.388	-0.048	1.612	0.966	0.825	1.869	1.701	1.737
Belgium	0.721	0.306	0.080	1.123	0.952	1.000	0.864	1.021	1.008
China	0.552	-0.074	-0.066	0.482	0.237	0.270	0.717	0.601	0.559
Croatia	0.838	0.376	-0.328	0.727	0.549	0.067	1.114	0.903	0.232
Czech Republic	1.032	0.320	-0.111	1.264	0.973	0.661	1.783	1.753	1.314
Denmark	0.955	0.199	0.085	0.977	0.630	0.557	1.009	0.725	0.577
Estonia	1.052	0.123	-0.201	1.401	0.878	0.036	1.349	0.820	0.008
Finland	0.763	0.437	0.175	1.090	1.292	0.997	1.137	1.323	1.041
France	0.677	0.278	0.062	1.184	1.022	0.910	1.120	1.184	1.066
Germany	0.763	0.350	0.139	1.120	1.053	0.958	0.973	1.046	0.913
Greece	0.865	0.123	-0.238	0.908	0.516	0.368	0.875	0.541	0.385
Hong Kong	0.825	0.409	0.234	1.120	1.004	0.962	1.426	1.718	1.991
Hungary	0.715	0.137	-0.239	1.047	0.362	0.233	1.126	0.821	0.630
India	0.643	0.198	-0.024	0.558	0.350	0.181	0.710	0.349	0.361
Ireland	0.546	0.253	0.160	0.824	0.681	0.615	0.840	0.683	0.589
Italy	0.862	0.258	-0.021	1.350	0.901	0.751	1.187	0.999	0.850
Jamaica	0.363	-0.013	-0.131	0.274	-0.013	-0.010	0.307	-0.076	-0.088
Japan	0.705	0.131	0.120	1.056	0.389	0.425	1.537	1.707	1.493
Luxembourg	0.931	0.174	0.008	1.502	1.015	0.849	1.185	0.905	0.932
Malaysia	0.652	0.548	0.735	0.803	0.648	0.850	1.447	1.703	1.730
Malta	0.775	0.103	-0.079	0.523	0.252	0.194	0.381	0.098	0.084
Mexico	0.549	0.128	0.077	0.646	0.613	0.560	0.807	0.828	0.950
Netherlands	0.740	0.266	0.048	1.297	1.157	1.151	0.998	1.284	1.312
Norway	1.396	0.379	-0.292	1.734	1.220	0.957	3.793	3.526	3.906
Poland	1.013	0.185	-0.195	1.229	0.831	0.568	1.614	1.209	0.939
Portugal	0.812	0.204	-0.061	1.207	0.810	0.752	0.967	0.823	0.685
Singapore	0.609	0.132	-0.142	0.871	0.624	0.491	1.246	1.726	1.394
Slovakia	0.752	0.224	0.033	0.564	0.261	0.066	0.440	0.104	-0.025
South Korea	0.514	0.170	-0.104	0.647	0.852	0.847	1.272	1.560	1.410
Spain	0.676	0.237	-0.046	1.191	0.926	0.757	0.890	0.880	0.718
Sweden	0.307	0.041	-0.051	0.780	0.404	0.316	0.846	0.697	0.593
Switzerland	0.690	0.362	0.159	0.976	0.809	0.678	0.714	0.783	0.724
Taiwan	0.626	0.428	0.241	0.905	0.888	0.920	0.998	1.483	1.372
Thailand	0.671	0.248	0.116	0.927	0.702	0.386	1.333	1.340	0.980
Turkey	0.565	0.138	0.176	0.744	0.450	0.511	0.850	0.029	0.011
United Kingdom	0.613	0.220	-0.041	1.250	0.981	0.857	1.192	1.406	1.459
United States	0.457	-0.047	0.000	0.481	0.480	0.570	0.579	0.177	0.341

Table A3: In-Sample R^2, and Out-of-Sample R^2 and Pooled out-of-sample R^2 results for three individual commodities (continued).

Country	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)
	Panel G. Gold			Panel H. Heating Oil			Panel I. Natural Gas		
Argentina	0.548	-0.018	-0.081	0.595	0.140	0.215	0.717	0.059	0.155
Australia	1.408	0.503	0.383	2.022	2.325	1.902	0.494	-0.030	-0.346
Austria	0.614	-0.077	-0.029	1.309	0.977	1.082	0.499	-0.055	-0.259
Belgium	0.423	-0.182	-0.093	0.614	0.445	0.454	0.275	-0.052	-0.074
China	0.355	-0.087	-0.061	0.581	0.056	0.082	0.360	-0.053	-0.023
Croatia	0.746	0.222	0.356	0.772	0.551	0.260	0.430	-0.224	-0.538
Czech Republic	1.145	-0.272	-0.025	1.332	1.180	0.910	0.542	-0.059	-0.232
Denmark	0.430	-0.136	-0.069	0.719	0.293	0.197	0.346	0.075	0.082
Estonia	1.226	0.364	0.087	0.942	0.452	0.075	0.399	-0.047	-0.249
Finland	0.356	-0.350	-0.387	0.732	0.464	0.551	0.327	-0.041	-0.062
France	0.411	-0.370	-0.308	0.754	0.542	0.461	0.259	-0.046	-0.108
Germany	0.417	-0.289	-0.353	0.653	0.376	0.325	0.263	-0.044	-0.118
Greece	0.625	-0.224	-0.059	0.679	0.393	0.255	0.295	-0.081	-0.070
Hong Kong	0.832	0.018	-0.019	1.223	1.258	1.423	0.303	-0.065	-0.182
Hungary	0.752	-0.309	-0.415	1.050	0.424	0.267	0.314	0.007	-0.155
India	0.469	0.075	0.111	0.556	0.186	0.187	0.600	-0.063	-0.441
Ireland	0.517	-0.333	-0.311	0.572	0.364	0.234	0.412	0.017	-0.074
Italy	0.468	-0.345	-0.163	0.894	0.392	0.278	0.387	-0.046	-0.078
Jamaica	0.284	-0.028	-0.072	0.273	-0.103	-0.131	0.270	-0.008	0.036
Japan	0.475	-0.070	-0.041	1.160	1.159	1.036	0.255	-0.055	-0.035
Luxembourg	0.532	-0.265	-0.187	0.764	0.438	0.474	0.354	-0.070	-0.170
Malaysia	0.888	0.581	0.708	1.342	1.302	1.155	0.562	-0.007	-0.234
Malta	0.809	0.338	0.513	0.297	-0.091	-0.095	0.380	-0.061	-0.066
Mexico	0.411	-0.159	-0.030	0.568	0.478	0.502	0.407	0.012	-0.005
Netherlands	0.388	-0.424	-0.367	0.707	0.600	0.646	0.320	-0.025	-0.097
Norway	0.769	-0.121	0.009	2.811	2.153	2.517	0.702	0.085	-0.189
Poland	1.010	-0.236	-0.076	1.284	0.721	0.546	0.350	-0.075	-0.244
Portugal	0.619	0.041	0.089	0.673	0.510	0.483	0.378	-0.004	-0.065
Singapore	0.655	0.283	0.245	1.165	1.266	0.857	0.570	0.075	-0.359
Slovakia	2.274	1.251	1.591	0.270	-0.032	-0.107	0.253	-0.189	-0.403
South Korea	0.877	0.044	0.051	1.041	1.154	1.054	0.324	-0.140	-0.230
Spain	0.509	-0.382	-0.313	0.670	0.374	0.287	0.286	-0.047	-0.134
Sweden	0.571	-0.306	-0.427	0.534	0.194	0.209	0.354	0.021	0.009
Switzerland	0.547	0.061	0.158	0.510	0.443	0.328	0.316	0.002	-0.083
Taiwan	0.781	0.622	0.406	0.920	1.187	1.106	0.307	0.000	-0.078
Thailand	1.097	0.571	0.619	1.260	1.011	0.802	0.329	-0.057	-0.060
Turkey	0.518	-0.096	-0.039	0.641	0.052	0.143	0.474	-0.049	-0.036
United Kingdom	0.367	-0.297	-0.048	0.963	0.902	0.910	0.339	0.039	-0.045
United States	0.283	-0.184	-0.198	0.545	0.156	0.239	0.253	-0.048	-0.110

Table A3: In-Sample R^2, and Out-of-Sample R^2 and Pooled out-of-sample R^2 results for three individual commodities (continued).

Country	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)
	Panel J. Silver			Panel K. Sugar			Panel L. Wheat		
Argentina	0.460	-0.133	-0.291	0.741	0.205	0.178	0.528	-0.212	-0.052
Australia	0.347	-0.054	-0.133	1.400	1.323	1.255	1.215	0.440	0.014
Austria	0.406	-0.038	-0.137	0.762	0.556	0.471	0.854	0.217	-0.104
Belgium	0.256	-0.129	-0.198	0.578	0.338	0.230	0.519	0.075	-0.110
China	0.359	0.015	0.014	0.532	0.211	0.118	0.770	-0.112	-0.220
Croatia	0.268	-0.251	-0.267	0.461	0.219	-0.017	0.557	0.102	-0.415
Czech Republic	0.451	0.001	-0.058	0.594	0.273	0.175	0.829	0.280	0.000
Denmark	0.336	-0.062	-0.087	0.551	0.128	0.174	0.763	0.218	0.020
Estonia	0.484	-0.084	-0.245	0.841	0.421	0.219	0.543	0.011	-0.241
Finland	0.640	-0.198	-0.323	0.740	-0.117	-0.303	0.536	0.098	-0.118
France	0.370	-0.089	-0.130	0.571	0.251	0.148	0.432	0.032	-0.159
Germany	0.363	-0.119	-0.175	0.483	0.160	0.029	0.478	0.052	-0.152
Greece	0.421	-0.134	-0.185	0.524	0.065	-0.103	0.590	0.147	-0.042
Hong Kong	0.477	-0.126	-0.194	0.759	0.612	0.583	0.755	0.319	0.281
Hungary	0.431	-0.147	-0.298	0.530	0.059	-0.026	0.697	0.136	-0.080
India	0.339	-0.278	-0.279	0.557	0.184	0.068	0.604	0.024	-0.157
Ireland	0.446	-0.083	-0.142	0.337	-0.081	-0.123	0.426	-0.062	-0.115
Italy	0.359	-0.060	-0.010	0.559	0.238	0.220	0.462	0.030	-0.113
Jamaica	0.211	-0.098	-0.161	0.191	0.021	-0.024	0.178	-0.102	-0.088
Japan	0.502	-0.022	0.034	0.834	0.308	0.220	0.650	0.209	0.144
Luxembourg	0.273	-0.108	-0.153	0.787	0.344	0.294	0.765	0.170	-0.009
Malaysia	0.244	-0.042	-0.157	1.067	0.649	0.848	0.296	0.085	0.116
Malta	0.332	-0.152	-0.157	0.661	0.042	-0.058	0.354	0.010	-0.101
Mexico	0.414	-0.081	-0.050	0.420	0.209	0.219	0.309	-0.093	-0.072
Netherlands	0.289	-0.117	-0.179	0.463	0.185	0.182	0.516	0.006	-0.298
Norway	0.368	-0.204	-0.236	0.929	0.654	0.629	0.825	0.232	-0.093
Poland	0.333	-0.078	-0.108	0.654	0.220	0.074	0.917	0.102	-0.151
Portugal	0.255	-0.089	-0.124	0.552	0.289	0.182	0.718	0.154	-0.062
Singapore	0.283	-0.066	-0.064	0.624	0.205	0.185	0.497	-0.229	-0.346
Slovakia	0.202	0.006	-0.090	0.371	0.038	-0.060	0.549	0.074	0.008
South Korea	0.549	-0.006	-0.073	0.670	0.450	0.308	0.577	0.050	-0.274
Spain	0.302	-0.119	-0.138	0.495	0.199	0.148	0.508	0.069	-0.166
Sweden	0.381	-0.038	-0.029	0.353	-0.128	-0.158	0.392	-0.130	-0.109
Switzerland	0.247	-0.038	-0.063	0.477	0.228	0.205	0.533	0.095	-0.136
Taiwan	0.357	-0.059	-0.076	0.679	0.743	0.462	0.536	0.090	-0.221
Thailand	0.300	-0.200	-0.222	0.692	0.549	0.290	0.516	0.073	0.101
Turkey	0.442	-0.129	-0.154	0.529	0.243	0.161	0.626	0.007	0.075
United Kingdom	0.294	-0.164	-0.269	0.491	0.300	0.224	0.410	-0.007	-0.142
United States	0.408	-0.023	-0.164	0.314	-0.052	-0.119	0.285	-0.138	-0.223

Table A4: In-Sample, Out-of-Sample and pooled R^2 Results for commodity indices. The first column indicates the country, followed by 3 sets of columns for each commodity panel. For each panel, the table reports the in-sample R^2 in percentage, the out-of-sample R^2 in percentage based on individual time-series estimates, and the out-of-sample R^2 in percentage based on pooled estimates.

Country	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)
	Panel A. Agriculture			Panel B. Energy			Panel C. Industrial Metals		
Argentina	0.841	0.025	0.300	0.847	0.180	0.256	0.950	0.183	0.784
Australia	3.035	1.777	0.976	2.846	3.491	2.987	5.398	5.289	4.549
Austria	1.688	0.809	0.205	1.706	1.669	1.889	3.371	2.212	1.929
Belgium	1.084	0.663	0.189	0.854	1.008	1.111	1.529	1.112	1.055
China	0.925	0.084	0.038	0.641	0.439	0.491	1.615	1.395	1.604
Croatia	0.928	0.535	-0.064	1.146	0.903	0.163	1.754	1.400	1.086
Czech Republic	1.366	0.762	0.389	1.780	1.621	1.301	3.072	2.601	2.060
Denmark	1.257	0.394	0.151	0.885	0.872	0.806	1.748	0.950	0.700
Estonia	1.420	0.705	0.014	1.244	0.889	0.072	2.471	1.742	1.020
Finland	0.997	0.657	0.001	1.126	1.296	1.287	1.682	1.629	1.622
France	0.966	0.552	0.070	1.109	1.160	1.193	1.522	1.235	1.149
Germany	1.020	0.570	0.064	0.981	1.006	0.869	1.534	1.238	1.252
Greece	1.013	0.433	-0.058	0.785	0.508	0.321	1.931	1.015	0.957
Hong Kong	1.481	1.019	0.645	1.469	1.822	2.034	2.600	2.973	2.914
Hungary	1.114	0.378	0.002	1.231	0.857	0.639	2.147	1.281	0.831
India	0.856	0.033	-0.109	0.751	0.451	0.400	1.085	0.978	0.931
Ireland	0.660	0.383	0.185	0.700	0.685	0.679	1.287	0.998	0.976
Italy	1.020	0.478	0.124	1.238	1.049	0.941	1.524	0.970	0.819
Jamaica	0.341	-0.031	-0.088	0.348	-0.058	-0.160	0.284	0.027	-0.008
Japan	1.182	0.410	0.290	1.413	1.608	1.485	2.960	2.216	2.019
Luxembourg	1.664	0.868	0.552	1.124	0.938	0.966	2.441	2.011	1.957
Malaysia	1.075	0.877	1.097	1.578	1.699	1.654	1.587	1.689	2.206
Malta	0.754	0.194	0.057	0.284	0.019	0.002	1.160	0.461	0.328
Mexico	0.607	0.274	0.168	0.923	0.893	0.976	0.906	0.855	1.017
Netherlands	1.084	0.507	-0.117	1.037	1.277	1.353	1.494	1.270	1.287
Norway	2.031	0.838	0.223	3.460	3.546	4.060	2.988	1.684	1.785
Poland	1.485	0.625	0.225	1.469	1.264	1.053	2.878	1.996	1.927
Portugal	1.234	0.624	0.248	0.998	0.957	0.869	1.686	1.130	0.949
Singapore	0.948	0.614	0.211	1.354	1.736	1.302	3.173	3.363	2.901
Slovakia	0.845	0.379	0.223	0.335	0.013	-0.098	1.144	0.772	0.671
South Korea	1.308	0.848	0.114	1.245	1.725	1.700	2.357	3.341	2.875
Spain	0.929	0.509	0.101	0.960	0.854	0.741	1.468	1.175	0.910
Sweden	0.571	0.129	-0.071	0.795	0.693	0.652	0.882	0.719	0.662
Switzerland	0.889	0.471	0.051	0.638	0.729	0.630	1.153	1.036	1.006
Taiwan	1.212	1.214	0.478	0.942	1.546	1.351	1.889	2.891	2.647
Thailand	0.984	0.637	0.502	0.971	1.118	0.898	1.364	1.259	1.346
Turkey	0.700	0.139	0.252	0.940	-0.061	-0.089	0.916	0.031	0.122
United Kingdom	1.008	0.580	0.041	1.222	1.475	1.631	1.538	1.369	1.338
United States	0.594	0.055	-0.071	0.606	0.248	0.371	0.669	0.390	0.678

Table A4: Commodity Indices In-Sample and Out-of-Sample R^2 Results (continued).

Country	In-Sample (%)	Out-Of-Sample (%)	Pooled (%)
	Panel D. Precious Metals		
Argentina	0.604	0.014	-0.014
Australia	1.612	0.728	0.602
Austria	0.762	-0.022	-0.103
Belgium	0.428	-0.152	-0.149
China	0.311	-0.115	-0.077
Croatia	0.683	0.296	0.297
Czech Republic	1.261	-0.043	-0.021
Denmark	0.461	-0.121	-0.125
Estonia	1.373	0.222	-0.207
Finland	0.343	-0.215	-0.171
France	0.409	-0.293	-0.319
Germany	0.474	-0.226	-0.331
Greece	0.698	-0.054	0.086
Hong Kong	0.965	0.148	0.071
Hungary	0.894	-0.196	-0.414
India	0.720	0.185	0.156
Ireland	0.552	-0.336	-0.457
Italy	0.453	-0.325	-0.158
Jamaica	0.340	-0.051	-0.115
Japan	0.476	-0.059	0.000
Luxembourg	0.571	-0.237	-0.200
Malaysia	0.982	0.766	0.975
Malta	0.826	0.339	0.449
Mexico	0.445	-0.180	-0.041
Netherlands	0.432	-0.385	-0.415
Norway	0.957	-0.107	0.008
Poland	1.287	-0.013	0.109
Portugal	0.681	0.060	0.098
Singapore	0.861	0.541	0.398
Slovakia	2.337	1.289	1.554
South Korea	0.930	0.193	0.373
Spain	0.500	-0.274	-0.287
Sweden	0.581	-0.278	-0.418
Switzerland	0.523	0.047	0.032
Taiwan	0.785	0.738	0.609
Thailand	1.228	0.663	0.721
Turkey	0.445	-0.053	-0.036
United Kingdom	0.360	-0.215	-0.021
United States	0.301	-0.147	-0.148

Table A5: The first 5 trading days are illustrated, where a long- or short position is taken based on a given signal. The three indices with the largest expected change (%) generate signals to take a long position, whereas the 3 lowest (or most negative) changes generate signals to take a short position in the index. R^2 values are reported for each index in the rightmost column.

Date	Index	Signal	Expected Change (%)	R²
2024-01-02	Czech Republic	Long	0.608	0.278
2024-01-02	Hungary	Short	0.603	0.345
2024-01-02	Turkey	Long	0.580	0.466
2024-01-02	Turkey	Short	-0.188	0.230
2024-01-02	Singapore	Short	-0.099	0.221
2024-01-02	Luxembourg	Short	-0.099	0.263
2024-01-03	Argentina	Long	1.240	0.489
2024-01-03	Turkey	Long	0.805	0.466
2024-01-03	Hungary	Long	0.214	0.348
2024-01-03	Norway	Short	-0.683	0.317
2024-01-03	Mexico	Short	-0.554	0.263
2024-01-03	Finland	Short	-0.491	0.301
2024-01-04	Turkey	Long	0.751	0.466
2024-01-04	Norway	Long	0.551	0.317
2024-01-04	Sweden	Long	0.432	0.227
2024-01-04	India	Short	-0.157	0.273
2024-01-04	Greece	Short	-0.152	0.367
2024-01-04	China	Short	-0.096	0.230
2024-01-05	South Korea	Long	0.274	0.329
2024-01-05	Japan	Long	0.104	0.182
2024-01-05	Jamaica	Long	0.064	0.091
2024-01-05	Argentina	Short	-1.007	0.489
2024-01-05	Mexico	Short	-0.809	0.263
2024-01-05	Finland	Short	-0.712	0.310

Table A6. The returns of the first month of trading are presented, calculated as the net return of the six daily positions.

Date	Return ($)
2024-01-02	352.32
2024-01-03	543.23
2024-01-04	147.38
2024-01-05	-789.83
2024-01-08	-428.60
2024-01-09	-305.95
2024-01-10	9.68
2024-01-11	-423.57
2024-01-12	-97.22
2024-01-16	183.37
2024-01-17	128.09
2024-01-18	171.14
2024-01-19	-755.75
2024-01-22	-260.50
2024-01-23	129.54
2024-01-24	-240.44
2024-01-25	-248.96
2024-01-26	311.94
2024-01-29	145.98
2024-01-30	-245.50
2024-01-31	107.20
2024-02-01	520.63

Table A7. Production and Consumption of Copper, Cotton, and Crude Oil of the year 2023. Panel A shows mine production, and estimated consumption of 2023 in 1000 metric tons. Panel B shows production and consumption for 2023 of cotton in 1000 metric tons. Panel C shows crude oil production and consumption for 2023 in 1000 barrels per day.

Panel A. Copper		
Country	Mine Production (2023)	Estimated Consumption (2023)
	In 1,000 metric tons	
Australia	778	556,00
Hong Kong	0	34
Japan	0	2 423,00
South Korea	0	657,00
Singapore	0	176
United States	1 130	3 808,00
Panel B. Cotton		
Country	Production (2023)	Consumption (2023)
	In 1,000 metric tons	
Argentina	338	n/a
Australia	1090	n/a
Mexico	196	n/a
South Korea	0.218	n/a
Turkey	698	n/a
United States	2630	n/a
Panel C. Crude Oil		
Country	Production (2023)	Consumption (2023)
	In 1,000 barrels/day	
Australia	383	1056
Hong Kong	0	268
Malaysia	565	930
Mexico	2 040	1 962
Norway	2 022	213
Singapore	0	1359
United States	19 358	18 984

Source: Data retrieved from U.S. Department of Agriculture, Foreign Agricultural Service (n.d.), U.S. Geological Survey (2025), and World Population Review (n.d.)

Figure A1: Out-of-Sample RMSE against in-sample training length in years. The y-axis is the out-of-sample RMSE, and the x-axis represents the in-sample training length in years. The dots represent each year and show the trend of the out-of-sample RMSE as the more data is included in the sample.