Forecasting Commodity Returns: A Comparative Study of Linear Machine Learning Models' Predictive Performance

MASTER'S DEGREE IN ECONOMICS AND FINANCE

Chair of Financial Economics

Paola Rizzitano

Supervisor: Nicola Borri
Co-Supervisor: Megha Patnaik
Student ID: 776591

Academic Year 2024/2025

ὁ δὲ ἀνεξέταστος βίος οὐ βιωτὸς ἀνθρώπῳ

— Plátōn, *Apología Sōkrátous*, 38a

Abstract

In a market context in which participants and policy makers increasingly demand real time forecasts for asset prices, this paper aims at identifying indicator variables and machine learning techniques that are most effective in predicting future commodity returns.

The emphasis on commodity returns rather than futures or spot prices is deliberate. By focusing on returns, the analysis abstracts from contract-specific features of futures markets—such as maturity effects and rollover dynamics—and instead adopts an investor-oriented perspective. In doing so, the paper treats commodities as a speculative asset class, reflecting their increasing use by institutional investors for return-seeking purposes rather than solely as hedging instruments.

On the selection of regressors for the forecasting exercise, the study combines economic and technical indicators in order to exploit both macroeconomic and market based information, and examines their predictive relevance within a linear regularization-based machine learning framework. More specifically, using monthly log-returns for three major commodities over a ten-year sample period, a progressive framework is adopted: starting from purely autoregressive models, the analysis moves to multivariate and regularized multivariate specifications, whose predictive accuracy is evaluated against traditional benchmark models.

Overall, these regularized techniques deliver a significant improvement in out-of-sample forecasting accuracy for commodity market returns, which are notoriously difficult to predict, highlighting the value of dimension reduction and/or variable selection in higher-noise financial environments.

Contents

Introduction

Commodity-assets' properties still remain quite unknown in the eyes of the average market participant, despite having being traded -mainly via futures- for more than a century. Unlike other financial assets, commodities tend to exhibit pronounced seasonality in price levels and volatility, but have recently emerged as a popular ”equity-like” asset class for many institutional investors. This phenomenon is generally attributed to the so-called financialization of commodity markets, proposed by Gorton et al. (2013) and Tang and Xion (2012), and recognizes a speculative purpose also for this asset class. Therefore, the increasing interest from market participants in effectively predicting the future behavior of commodity returns should not come as a surprise.

Moreover, being aware of the importance of understanding and forecasting commodity returns for both academic and practical relevance, this thesis aims at investigating predictive power of combining macro-financial and technical indicators for three key commodities - gold, corn, and natural gas- and evaluates their performance through traditional multivariate least-square regressions as well as modern regularization techniques, such as LASSO and Ridge. By comparing these frameworks against Random Walk models, the analysis aims to provide a comprehensive assessment of which signals are genuinely informative for commodity return prediction and how model choice affects out-of-sample forecasting performance.

More specifically, building on Gargano and Timmermann (2014), the set of regressor variables is composed of macroeconomic and financial indicators, and by merging technical variables proposed by Wang et al. (2020), the purpose of the first part of the study is to assess whether certain signals consistently enhance forecasting performance across different commodities, or whether their relevance is largely asset-specific.

The analysis then shifts to the core objective of the paper, this being to examine how different modeling approaches influence predictive accuracy, with a particular focus on the incremental benefits of modern regularization techniques. As a matter of fact, while Ordinary Least Square (OLS) provides an intuitive and straightforward framework, it can be sensitive to multicollinearity and prone to overfitting when the number of predictors is large. In contrast, LASSO and Ridge regression enforce sparsity and shrink coefficients, allowing for a systematic evaluation of whether a regularized set can improve out of sample forecasting performance.

Lastly, results are compared to more naive benchmarks, these being the Random Walk with and without drift models, that are usually considered as a baseline forecast in commodity markets, providing a natural reference point for assessing the added value of more sophisticated approaches.

Overall, this paper aims at contributing to the existing literature in several ways. First, the focus of the thesis departs from common commodity forecasting related literature by aiming at the determination of log returns rather than futures prices. This difference is crucial for understanding investment purposes, even though results are directly linked to futures prices properties. When trying to predict futures prices, as stated by Gorton and Rouwenhorst (2006), it is important to keep in mind several aspects, among which: (1) commodity futures are derivative securities, and therefore do not represent claims on corporations; and (2) commodity futures are short-maturity claims on real assets, implying that to maintain the position through time the investor needs to roll it at every expiration date. The key takeaway is that returns from holding the position depend on price levels of the futures contracts, tending to compensate investors willing to bear the risk of short term price fluctuations: focusing on returns rather than price levels naturally aligns the analysis to an investment-based perspective, emphasizing the role of commodities as financial assets and allowing for a direct assessment of their speculative and risk-return properties.

As a second contribution, this thesis provides an empirical assessment of the use of traditional economic variables and technical indicators in forecasting models, clarifying which variables are most informative for the selected dataset, and measures the different impacts according to the specific commodity class. Building on Gargano and Timmermann (2014), the recognition of the importance of including macroeconomic and financial variables is inherited, as predictability of commodity prices can be expected to be driven by time-varying storage costs and convenience yields: these, in fact, are believed to be influenced by the state of the economy and financing costs. However, departing from the traditional approach, also technical indicators are included in the multivariate model, as proposed by Wang et al. (2020). Technical rules aim at identifying future price trends according to past prices or volume patters, but have received less attention in the literature for forecasting commodity returns. Given that this paper recognizes the increasingly speculative and financialized nature of commodity markets, technical indicators may convey valuable information about traders' behavior and market sentiment, complementing macroeconomic and financial signals in explaining return dynamics.

Lastly, the paper evaluates the progressive benefits in adding layers of sophistication through regularization models, comparing OLS, LASSO, and Ridge in a unified out-of-sample framework, highlighting both predictive performance and variable selection over time. This perspective is particularly relevant for commodity returns, which are generally characterized by weak and unstable predictability. By keeping the underlying set of predictors constant across models, the analysis isolates the marginal contribution of each estimation method, allowing for a clearer assessment of whether and how regularization techniques improve forecasting accuracy beyond standard linear regressions.

The structure of the remainder of the thesis is as follows. Chapter 1: *Data Presentation and Analysis* aims at familiarizing the reader with the dataset of commodities, by showing main statistical features and visual time series representations. The same insights are given on the regressor variables that will be then used when implementing the forecasting exercise, with a focus on how technical indicators are built. Chapter 2: *Forecasting Methodology* serves at illustrating how the models work from a theoretical perspective and how they are implemented in the predictive exercise. Chapter 3: *Results* reports the out-of-sample forecasting results, starting with expanding window OLS, followed by LASSO and Ridge regressions, and compares their performance to Random Walk benchmarks. The chapter also examines which predictors are retained under regularization and interprets their economic relevance. Finally, Chapter 5: *Discussion* synthesizes the main findings and contextualizes them to the specific features of each commodity class, and then discusses practical implications for investors and portfolio managers. Elements for future research are also proposed in the conclusion of the paper.

1

Data Presentation and Analysis

1.1 Commodity Futures Prices

For the purpose and simplification of the thesis, three commodities are selected and studied, where each is assumed to be a representative of the commodity class to which it belongs. The choice of these three commodities does not aim at exhausting the heterogeneity of commodity markets, but rather at providing representative proxies for three major commodity classes, allowing for a meaningful comparison across categories.

Namely, *gold* is considered as a representative for the precious metals commodity class; *corn* for agricultural commodities; and *natural gas* is studied in relation to energy commodities.

The analysis focuses on predicting returns through monthly data obtained for each of the commodities' futures, preferring these to spot prices due to their forward looking nature: prices reflected by futures contracts embed expectations on future spot prices, and any positive return is generated when unexpected movements occur. For the sake of clarity, an investor with a long position in futures will benefit when the price of the underlying increases by more than expected by the market because the spot price will approach the futures' one at maturity, increasing the value of the position. Moreover, any difference between the expected futures price and the actual future price of the underlying is called the risk premium. It is important to note that speculators do not have to hold the contract until expiration to earn the premium, but they do need to roll over the position if they wish to extend their exposure over time. The use of continuous futures series implicitly accounts for this rollover mechanism, aligning the empirical analysis with the actual returns faced by market participants.

On the dataset, the span goes from January 2015 to November 2025, collecting approximately the last ten years of data. For a visual representation, Figures 2.1 through 2.6 plot the respective time series of prices of *gold*, *corn*, and *natural gas* for the mentioned period, accompanied by rolling mean and volatility representations.

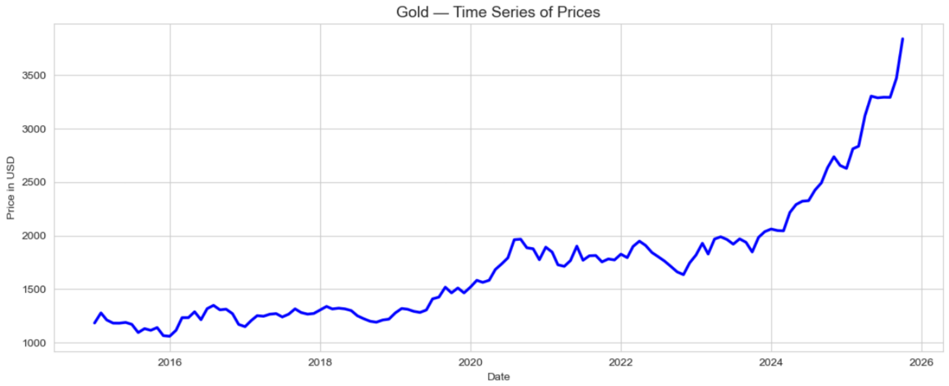

Figure 1.1: Time Series of Prices for *gold* from January 2015 to November 2025.

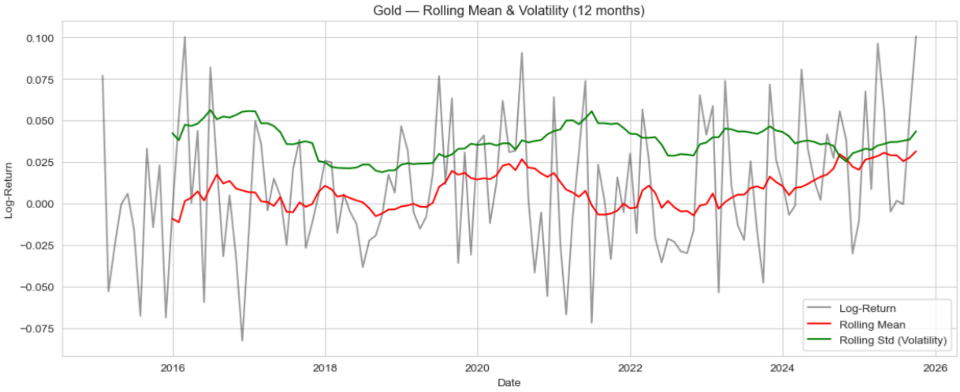

Figure 1.2: Rolling Mean and Volatility for *gold* from January 2015 to November 2025.

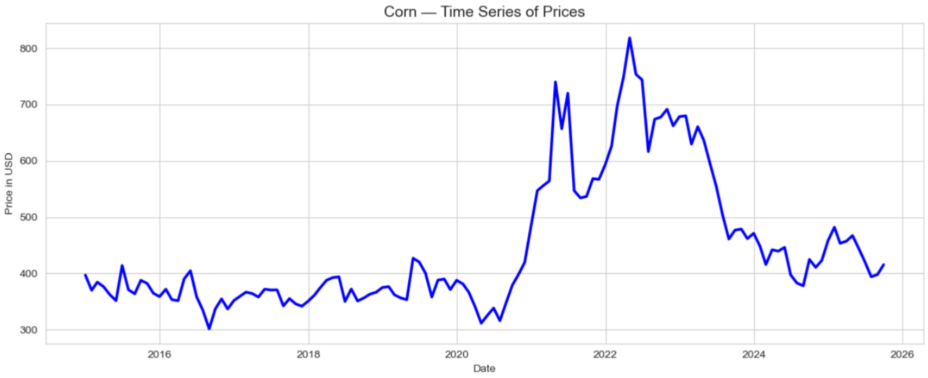

Figure 1.3: Time Series of Prices for *corn* from January 2015 to November 2025.

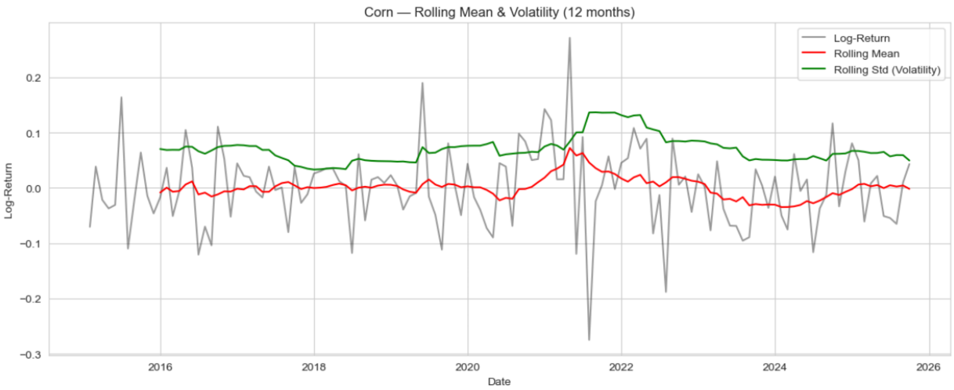

Figure 1.4: Rolling Mean and Volatility for *corn* from January 2015 to November 2025.

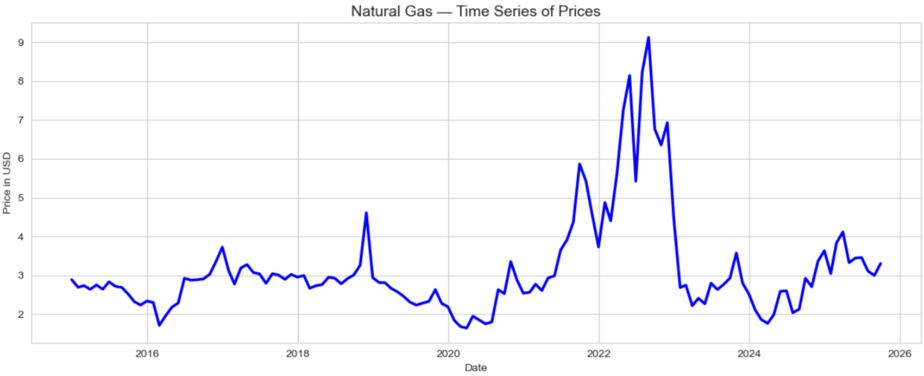

Figure 1.5: Time Series of Prices for *natural gas* from January 2015 to November 2025.

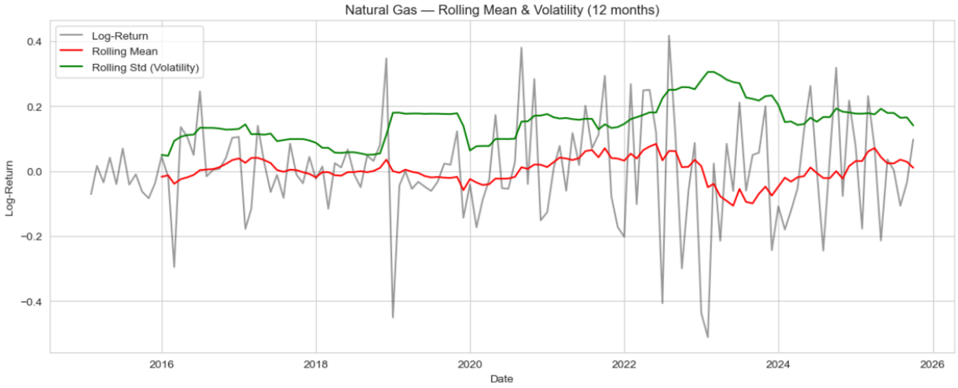

Figure 1.6: Rolling Mean and Volatility for *natural gas* from January 2015 to November 2025.

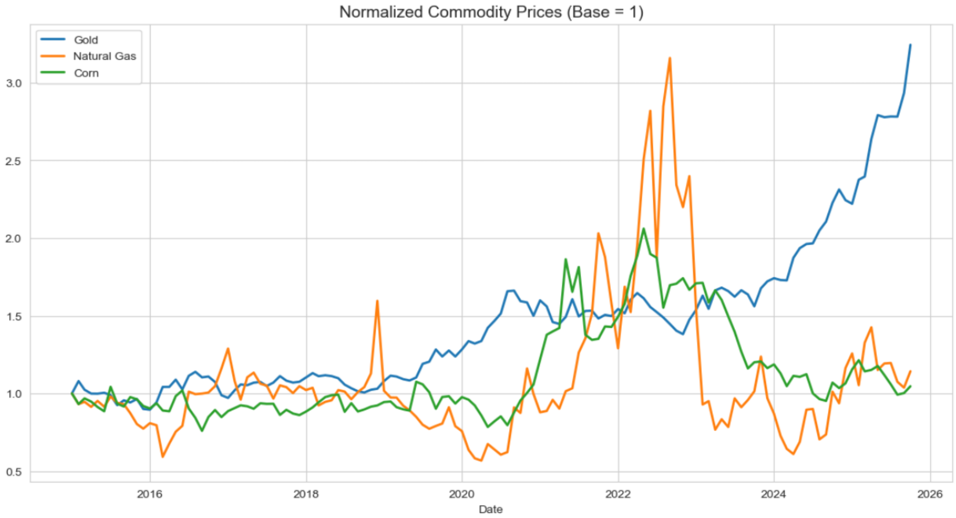

Figure 1.7: Combined Time Series of Commodity Prices for Visual Comparison from January 2015 to November 2025.

A preliminary visual inspection suggests a degree of co-movement across the three price series, together with pronounced volatility. Therefore, already at a first glance, two characteristics emerge: high intra-asset volatility and positive correlation across commodities.

For the purpose of this study, as stated above, commodity futures prices will be modeled in order to obtain logarithmic returns, our dependent variable. The equation can be written as follows:

$$r_{t+1} = \log(P_{t+1}) - \log(P_t),$$

where P_t denotes the price at time t, and r_t the log return at time t. The choice of using logarithmic returns is justified by the exploitable properties of normality of log returns, their stationarity, and their being additive over time, which is convenient for empirical analysis and forecasting exercises. Table 1.1 summarizes the properties of each commodity's returns.

Table 1.1: Summary Statistics for Commodity Returns

Commodity	Mean	St. Dev.	Skew.	Ex. Kurt.	Min	1%	99%	Max
Gold	0.0091	0.0397	0.1804	-0.3738	-0.025	-0.0708	0.0990	0.1005
Corn	0.0004	0.0723	0.0390	2.1916	-0.2748	-0.1692	0.1823	0.2712
Natural Gas	0.0010	0.1574	-0.2750	1.2905	-0.5112	-0.4467	0.3705	0.4168

Gold is the commodity that exhibits highest risk adjusted returns, immediately visible through the highest mean and lowest standard deviation levels of the set; positive skewness, implying that positive returns have occurred more often than negative ones; and negative kurtosis, signaling thinner tails relative to a normal distribution.

In contrast, Natural gas and Corn show lower risk adjusted returns, with the former also accounting for negative skewness, while the latter exhibiting fatter tails.

1.2 Macroeconomic and Financial Variables

To perform the regression analysis expected from this study, seven combined economic variables have been considered. Following Gargano and Timmermann (2014), these have been classified into two groups: macroeconomic and financial variables.

The first set comprehends monthly inflation levels and industrial production activity, both referring to the Federal Reserve Economic Data.

These macroeconomic variables are chosen due to their ability to strongly impact commodity prices' outlook. This is because inflation directly affects the value of commodities, that exhibit higher prices, and, as a matter of fact, are often chosen as investments for inflation hedging purposes. Industrial production should also be expected to have an upwards impact on the value of commodities, although a distinction could be made according to the nature of a commodity of being an input or output of the production process. Overall, we expect a positive impact for natural gas and corn, backed by the reason that an increase in the industrial production leads to a larger use of these input goods, therefore consuming inventories and increasing demand. The effect is ultimately visible through higher price levels. For gold, the outcome could be less straightforward, as its scope often departs from being an input good, tending towards luxurious or storage of value scopes.

The second set contains variables that reflect financial markets activities. Among these there is the three month (short term) treasury rate, the ten years (long term) yield, the excess return from the S&P 500 vs. the risk free interest rate , the volatility level recorded from the VIX[1], and exchange rate movements captured by the DXY [2].

The choice of including financial variables for forecasting commodity prices follows the acceptance of the theory of financialization of commodity markets promoted by Gorton et al. (2013). According to this theory, high volatility and increasing correlation between commodity and stock markets are signals of investors' allocation of speculative capital to this asset class. Therefore, over the years, commodities have been behaving more and more as financial assets, replicating an equity-like behaviour.

---

[1] The CBOE VIX calculates the implied volatility of the S&P 500 index options, and represents monthly expectations of stock market behavior.

[2] The DXY Index, or USD index, measures the strength of the US Dollar against a basket of currencies, where each is assigned a specific weight. The basket comprehends the Euro (approx. 57% weight), Yen, GBp, Canadian Dollar, Swedish Crown, and CHF.

Table 1.2 reports summary statistics for the macroeconomic and financial variables employed in the forecasting analysis. All variables are expressed at a monthly frequency and transformed to ensure stationarity.

Table 1.2: Summary Statistics for Macroeconomic Variables

Variable	Mean	Std. Dev.	Min	Max	Skewness	Kurtosis	Obs
T-Bill	0.0193	0.0191	0.0000	0.0532	0.6153	-1.1626	129
Long Term Yield	0.0260	0.0114	0.0054	0.0488	0.2440	-0.9656	129
Exchange Rate	4.5873	0.0496	4.4901	4.7196	0.3782	-0.3871	129
Excess Return	0.0076	0.0437	-0.1337	0.1193	-0.5511	0.6760	129
Volatility	2.8599	0.3232	2.2523	3.9804	0.7080	0.4053	129
Inflation	0.0025	0.0028	-0.0080	0.0129	0.1810	3.3097	129
Industrial Production	-0.0002	0.0160	-0.1414	0.0635	-4.9485	49.0017	129

1.3 Technical Indicators

Following Wang et al. (2020), a set of technical indicators is constructed and employed to perform forecasting of commodity returns. These indicators are based on simple trading rules and generate binary signals taking a value of one when a buy (sell) condition is met, and zero otherwise. The underlying idea is that price dynamics may embed exploitable patterns reflecting investors' behaviour and market momentum.

Momentum

The first technical indicator is Momentum (MOM). Momentum strategies are based on the idea of buying assets that have recently experienced price increases and selling those that have underperformed. Formally, the momentum signal is defined as:

$$S_{t,\mathrm{MOM}} = \begin{cases} 1, & \text{if } P_t \geq P_{t-k} \\ 0, & \text{otherwise} \end{cases}$$

where k is the look-back period, meaning that we compare the current price P_t with the price of the asset at $t-k$, with $k = 1, 3, 6, 9, 12$ months, yielding a total of five momentum signals for each commodity.

Filtering Rule

The Filtering Rule (FR) generates trading signals based on percentage price movements relative to recent extrema. A buy signal is issued whenever the price increases by more than a predefined threshold (one percent in our study), while a sell signal is triggered following a sufficiently large decline:

$$S_{t,\mathrm{FR}} = \{S^{\mathrm{buy}}_{t,\mathrm{FR}}, S^{\mathrm{sell}}_{t,\mathrm{FR}}\}$$

$$S^{\mathrm{buy}}_{t,\mathrm{FR}} = \begin{cases} 1, & \text{if } P_t \geq (1+\eta/100) \times \min(P_{t-1},\ldots,P_{t-k}) \\ 0, & \text{otherwise} \end{cases}$$

$$S^{\mathrm{sell}}_{t,\mathrm{FR}} = \begin{cases} 1, & \text{if } P_t \leq (1-\eta/100) \times \max(P_{t-1},\ldots,P_{t-k}) \\ 0, & \text{otherwise} \end{cases}$$

where $\eta = 1\%$ and $k = 1, 3, 6, 9, 12$.
This rule therefore produces ten indicators per commodity: five buy and five sell signals.

Moving Average

The Moving Average (MA) rule compares short-term and long-term price trends to smooth out high-frequency noise. A buy (sell) signal is generated when the short-term moving average exceeds (falls below) the long-term moving average:

$$S_{t,\mathrm{MA}} = \begin{cases} 1, & \text{if } MA_{s,t} \geq MA_{l,t} \\ 0, & \text{otherwise} \end{cases}$$

where the moving average of order $j \in \{s, l\}$ is defined as:

$$MA_{j,t} = \frac{1}{j} \sum_{i=0}^{j-1} P_{t-i}$$

This indicator captures medium-term trends while mitigating short-run price fluctuations.

Oscillator Trading Rule

The Oscillator Trading Rule (OSLT) aims to identify overbought or oversold market conditions by exploiting potential trend reversals. The rule is based on the Relative Strength Index (RSI):

$$RSI(m) = 100 * \frac{U_t(m)}{U_t(m) + D_t(m)}$$

where upward and downward price movements over the past k periods are given by:

$$U_t(k) = \sum_{j=0}^{k-1} I(P_{t-j} - P_{t-j-1} > 0)(P_{t-j} - P_{t-j-1})$$

$$D_t(k) = \sum_{j=0}^{k-1} I(P_{t-j} - P_{t-j-1} < 0)|P_{t-j} - P_{t-j-1}|$$

with $I(\cdot)$ denoting the indicator function, equal to one when the condition in parentheses is satisfied. The trading signal is then defined as:

$$S_{t,\mathrm{OSLT}} = \{S^{\mathrm{buy}}_{t,\mathrm{OSLT}}, S^{\mathrm{sell}}_{t,\mathrm{OSLT}}\}$$

$$S^{\mathrm{buy}}_{t,\mathrm{OSLT}} = \begin{cases} 1, & \text{if } RSI \leq 50 + \gamma \\ 0, & \text{otherwise} \end{cases}$$

$$S^{\mathrm{sell}}_{t,\mathrm{OSLT}} = \begin{cases} 1, & \text{if } RSI \geq 50 + \gamma \\ 0, & \text{otherwise} \end{cases}$$

where $k = 1, 3, 6, 9, 12$ and $\gamma = 5, 10$.
This rule yields twenty oscillator-based indicators per commodity.

Support and Resistance Rule

The Support and Resistance (SR) rule compares the current price to historical extrema. Buy and sell signals are defined as:

$$S_{t,\mathrm{SR}} = \{S^{\mathrm{buy}}_{t,\mathrm{SR}}, S^{\mathrm{sell}}_{t,\mathrm{SR}}\}$$

$$S^{\mathrm{buy}}_{t,\mathrm{SR}} = \begin{cases} 1, & \text{if } P_t \geq (1 + \gamma/100) \times \max(P_{t-1}, \ldots, P_{t-k}) \\ 0, & \text{otherwise} \end{cases}$$

$$S^{\mathrm{sell}}_{t,\mathrm{SR}} = \begin{cases} 1, & \text{if } P_t \leq (1 - \gamma/100) \times \min(P_{t-1}, \ldots, P_{t-k}) \\ 0, & \text{otherwise} \end{cases}$$

with $k = 1, 3, 6, 9, 12$ and $\gamma = 1, 2, 3, 4, 5$.

Combining all five technical rules results in a 128×90 signal matrix for each commodity, where each column corresponds to a distinct trading signal and each row represents a monthly observation. In the next section, these technical indicators are aggregated and incorporated into the forecasting framework.

2

Forecasting Methodology

2.1 Autoregressive Models

As a first step for the forecasting analysis, commodity returns are forecasted using autoregressive models (AR). These models represent one of the most widely adopted benchmarks when it comes to time series forecasting, and are particularly suitable for financial return series, which might embody short term dependence but limited long run predictability.

An autoregressive framework of order ρ is denoted as AR(ρ), and assumes that a return at time $t+1$ depends linearly on its own past realizations:

$$r_{t+1} = \alpha + \sum_{j=1}^{\rho} \phi_j r_{t+1-j} + \varepsilon_{t+1}, \tag{2.1}$$

where r_{t+1} denotes the log return of the commodity one time forward, α is a constant, ϕ_j are autoregressive coefficients, and ε_{t+1} is an innovation term with mean zero and finite variance.

This type of specification aims to capture the short term persistence or mean reversion effects, while maintaining a parsimonious structure. In this study, autoregressive models are adopted to give a baseline forecasting tool rather than a structural model, before progressing towards more sophisticated techniques.

In line with standard practice in forecasting literature, an expanding window approach is used in order to estimate the models. This means that at each t, known as the forecasting point, the model is re-estimated using all the information available up to t, and a one-step-ahead forecast, $t+1$, is produced.

More formally, for each out-of-sample period, parameters are obtained by estimating an AR(5) model on observations $\{r_1, \ldots, r_t\}$ and the forecast $\hat{r}_{t+1}$ is computed using the estimated coefficients. For the sake of precision, it must be noted that the AR(5) model is estimated via OLS, by treating the lagged returns as regressors. This procedure ensures that only information available until t is used, thereby avoiding any lookahead bias. Moreover, using an expanding window in the commodity space appears to be better suited since structural breaks or evolving dynamics may occur over time. Re-estimating the model at each step allows to gradually incorporate and adapt to changing and additional information.

Autoregressive models have an important role in this study. Not only do they provide a purely time-series based benchmark that does not rely on exogenous predictors, allowing therefore to assess the incremental value added by macroeconomic and technical variables, but they also offer a natural comparison with the random walk hypothesis for prices, which implies that returns follow a martingale difference sequence with constant mean.

Under the random walk hypothesis for prices, the return process can be written as:

$$r_{t+1} = \mu + r_t + \varepsilon_{t+1}, \tag{2.2}$$

where μ represents the drift term. In empirical analysis, these benchmarks are implemented by using lagged returns as a forecast and employing an expanding mean of past returns as a drift component. The underlying notion conveyed by the random walk is that of market efficiency, according to which past prices have no predictive power for future ones. In this paper the same principle is applied focusing on predictive power of realized returns, rather than prices.

Moreover, in order to evaluate the forecasting performance of the autoregressive model, we consider the out-of-sample setting, as the sample is split according to an in-sample estimation period and out-of-sample estimation period. According to the literature, i.e., Campbell and Thompson (2007), forecast accuracy is estimated by calculating the Mean Squared Prediction Error (MSPE) and then using the percentage out-of-sample R^2_{Oos} to evaluate forecast performance. In this study we follow this approach and calculate:

$$R^2_{\text{Oos}} = 100 \times \left(1 - \frac{\text{MSPE}_{\text{model}}}{\text{MSPE}_{\text{benchmark}}}\right), \tag{2.3}$$

where

$$\text{MSPE}_i = \frac{1}{T-M} \sum_{t=M+1}^{T} \left(\hat{r}_i - r_t\right)^2, \quad i = \text{benchmark, model.} \tag{2.4}$$

To interpret the out-of-sample R^2_{Oos}, it is sufficient to study its sign: if positive, then the model outperforms the random walk benchmark, while negative values would suggest inferior performance.

2.2 Multivariate Ordinary Least Squares

After establishing the autoregressive model as a pure time-series benchmark, the forecasting analysis is extended by introducing predictive regressions that incorporate macroeconomic and technical indicators, following the approach of Wang et al. (2020). The aim of this part of the study is to assess whether external information beyond past returns can improve the predictability of commodity returns relative to the AR(5) and random walk models.

To this end, a dataset including both macroeconomic variables and technical trading signals is constructed. The former are intended to capture broad economic and financial conditions, while the latter serve as proxies for market participants' behavior and trading dynamics. Following the literature, i.e. Gargano and Timmermann (2014), the following predictive regression is implemented:

$$r_{t+1} = \alpha + \boldsymbol{\beta}^{\top}\mathbf{X}_t + \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim \mathcal{N}(0, \sigma^2), \tag{2.5}$$

where r_{t+1} denotes the commodity return, $\mathbf{X}_t$ is a vector of predictor variables, and ε_{t+1} is an independently and identically distributed disturbance term. The vector of coefficients $\boldsymbol{\beta}$ is estimated via Ordinary Least Squares (OLS).

Before proceeding with the forecasting exercise, an initial in-sample predictive regression is estimated to explore the relationship between commodity returns and the selected predictors. This step is purely diagnostic and exploratory, aimed at evaluating the sign, magnitude, and statistical significance of each predictor within a linear predictive framework. These results are not interpreted as evidence of genuine forecasting ability, as they do not account for parameter instability, model uncertainty, or information constraints.

Commodity markets are characterized by evolving dynamics, structural changes, and regime shifts. For this reason, conclusions regarding predictive performance are deferred to the out-of-sample analysis, in which model parameters are sequentially re-estimated using only information available up to each point in time.

The forecasting exercise begins by aligning predictors so that only information available at time t is used to forecast returns at time $t + 1$. The dependent variable is constructed as the lead of commodity log returns, while explanatory variables include the macroeconomic indicators and technical signals discussed above.

In line with standard forecasting practice, model estimation and evaluation are conducted on separate samples. Specifically, an expanding window approach is employed. The sample is split into an in-sample estimation period comprising 90% of the initial observations, while the remaining 10% forms the out-of-sample forecasting period.

At each time t, the predictive regression is re-estimated using all available information from the start of the sample up to t. The estimated coefficients are then used to generate a one-step-ahead forecast. This procedure is repeated sequentially until the end of the sample, producing a series of out-of-sample forecasts.

This expanding window approach is particularly well suited in this context, as it allows the model to incorporate new information over time while preserving estimation efficiency in relatively short samples. At the same time, coefficients are allowed to evolve gradually as additional data becomes available.

A key limitation of OLS-based predictive regressions is the risk of overfitting when a large number of explanatory variables is included. This concern is particularly relevant for technical indicators, as the technical trading rules described earlier generate a large set of highly correlated signals. Including each indicator separately would lead to a high-dimensional and potentially unstable forecasting environment.

To address this issue, and following Wang et al. (2020), a rule-based aggregation approach is adopted. Rather than selecting a single indicator for each trading rule, an equally weighted average is computed across all indicators belonging to the same rule, producing only one composite signal per trading strategy.

Formally, let $\hat{r}_{i,j,t}$ denote the forecasted commodity return generated by the i-th technical indicator associated with trading rule j, where $j \in \{\mathrm{MOM}, \mathrm{FR}, \mathrm{MA}, \mathrm{OSLT}, \mathrm{SR}\}$. The aggregated forecast is defined as:

$$\hat{r}_{j,t} = \frac{1}{P_j} \sum_{i=1}^{P_j} \hat{r}_{i,j,t}, \tag{2.6}$$

where P_j denotes the total number of technical indicators generated by trading rule j.

This aggregation step substantially reduces the dimensionality of the technical information set while preserving the core predictive content of each trading rule. It allows the model to avoid including dozens of highly correlated indicators, instead relying on a limited number of averaged technical signals, each representing a distinct trading philosophy.

From an econometric perspective, this approach mitigates multicollinearity and overfitting concerns, which are especially severe in small-sample forecasting settings. Moreover, averaging across indicators reduces sensitivity to specific parameter choices, enhancing robustness.

In the empirical implementation, five aggregated technical predictors are obtained and combined with the macroeconomic variables to form the final set of predictors used in the expanding window regressions. Forecasting performance is then evaluated using the Mean Squared Prediction Error (MSPE) and the corresponding out-of-sample R^2_{Oos}, comparing the model against the random walk benchmark.

Finally, statistical significance of out-of-sample predictability is assessed using the Clark and West (2007) test, which adjusts MSPE comparisons for nested models. The corrected loss differential is defined as:

$$L_t = (r_t - \bar{r}_t)^2 - (r_t - \hat{r}_t)^2 + (\bar{r}_t - \hat{r}_t)^2, \tag{2.7}$$

where the last term represents the adjustment relative to the Diebold and Mariano (1995) statistic. The null hypothesis is tested against the alternative:

$$\begin{aligned} H_0 &: \text{MSPE}_{\text{benchmark}} \leq \text{MSPE}_{\text{model}}, \\ H_1 &: \text{MSPE}_{\text{benchmark}} > \text{MSPE}_{\text{model}}, \end{aligned}$$

with statistical significance typically assessed at the 5% level.

2.3 Regularized Predictive Regressions: Lasso

This study acknowledges the limitations arising when the number of predictors is large within a multivariate OLS framework. The joint inclusion of several macroeconomic indicators and technical signals may lead to overfitting, multicollinearity, and parameter instability, ultimately deteriorating the out-of-sample forecasting performance of the model. These concerns are particularly relevant in commodity markets, where predictive relationships are known to be time-varying and subject to structural changes. For this reason, the forecasting analysis is extended by implementing the Least Absolute Shrinkage and Selection Operator (LASSO).

The adoption of this model allows for the simultaneous incorporation of multiple sources of information while departing from the approach proposed by Wang et al. (2020), where each predictor generates a separate forecast that is subsequently combined through a weighted average. In contrast, model uncertainty is addressed ex ante in this study: rather than estimating a large number of univariate models and aggregating their forecasts ex post, all macroeconomic variables and technical indicators are jointly included within a single predictive, yet regularized, regression framework.

Formally, LASSO extends the standard OLS objective function by introducing an ℓ_1 penalty on the regression coefficients, inducing both shrinkage and variable selection. The LASSO estimator solves the following optimization problem:

$$\min_{\boldsymbol{\beta}} \left\{ \sum_{t=1}^{T} \left(r_{t+1} - \alpha - \boldsymbol{\beta}^{\top} \mathbf{X}_t \right)^2 + \lambda \sum_j |\beta_j| \right\}, \tag{2.8}$$

where $\lambda \geq 0$ is the regularization parameter controlling the strength of the penalty. As λ increases, coefficients associated with weak or redundant predictors are progressively shrunk towards zero, allowing the model to retain only the most relevant variables at each point in time.

From an implementation perspective, the LASSO forecasting exercise closely mirrors the multivariate OLS setting described above. The same expanding window scheme is employed, with parameters estimated at each time t using only information available up to that date. One-step-ahead forecasts are generated sequentially until the end of the sample. All regressors are standardized within each estimation window to ensure appropriate penalization across variables measured on different scales. Moreover, the regularization parameter λ is selected via cross-validation at each estimation point.

Finally, forecasting accuracy is evaluated using the same MSPE-based metrics and the Clark and West (2007) test, allowing for a direct and fair comparison with the autoregressive and multivariate OLS models. Since the evaluation framework is held constant across specifications, any difference in forecasting performance can be attributed solely to the choice of the underlying model.

2.4 Regularized Predictive Regression: Ridge

Following the implementation of the multivariate OLS and LASSO models, this study proposes an additional regularization method aimed at shrinking less informative coefficients towards zero. The purpose of this step is to impose an ℓ_2 regularization that may improve robustness and stability of return predictability when a relatively large set of predictive variables is included.

Ridge regression estimates model parameters by minimizing a penalized least squares objective function:

$$\hat{\boldsymbol{\beta}}^{\text{ridge}} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{t=1}^{T} \left(r_{t+1} - \alpha - \boldsymbol{\beta}^{\top}\mathbf{X}_t \right)^2 + \lambda \sum_{j=1}^{K} \beta_j^2 \right\}, \tag{2.9}$$

where $\lambda \geq 0$ is the regularization parameter controlling the degree of shrinkage applied to the regression coefficients. Unlike LASSO regression, Ridge regression does not set any coefficient exactly equal to zero; instead, all coefficients are continuously shrunk toward zero as λ increases.

This feature is particularly relevant in the present context, where macroeconomic variables and technical indicators may be highly correlated. In such settings, excluding potentially informative predictors entirely may lead to a loss of relevant information and, consequently, to a deterioration in forecasting performance.

The Ridge model is implemented using the same expanding window forecasting framework adopted for the OLS and LASSO models. At each time t, the model is re-estimated using all available information up to that date, and a one-step-ahead forecast is generated. The regularization parameter λ is selected via cross-validation at each estimation step, allowing the degree of shrinkage to adapt dynamically over time as new information becomes available.

Importantly, while LASSO emphasizes sparsity and variable selection, Ridge regression focuses on coefficient stability. Comparing the two approaches provides insights into whether predictability is driven by a small subset of predictors or by a broader information set with diffuse predictive content.

Finally, once Ridge regression has been implemented for all commodities, out-of-sample forecasting performance relative to the benchmark models is evaluated using the MSPE and the corresponding out-of-sample R^2_{Oos} measure. Statistical significance is assessed using the Clark and West (2007) test.

3

Results

This section presents the main forecasting results obtained by combining macroeconomic variables and technical indicators within expanding regression frameworks. Predictions are generated over the 10% out-of-sample fraction of the ten-year period. The analysis aims to assess whether predictive regression models are able to outperform standard benchmark specifications in terms of forecasting accuracy.

Forecast performance is evaluated using the Mean Squared Prediction Error (MSPE) and formal statistical inference procedures. Results are reported separately for each commodity, allowing for a comparison of predictability across asset classes. The empirical findings are interpreted in light of the distinct characteristics of gold, corn, and natural gas markets, highlighting how differences in market structure, storage properties, and demand–supply dynamics may affect return predictability.

3.1 Autoregressive Models as a Time-Series Benchmark

As a preliminary benchmark, autoregressive models are estimated for all three commodities, focusing on the AR(5) specification and evaluating its out-of-sample performance relative to standard random walk benchmarks.

GOLD

The AR(5) predictive model is first applied to gold returns. The results highlight two main findings. First, the autoregressive model improves predictability relative to a pure random walk benchmark. Second, this improvement does not hold when the comparison is made against a random walk with drift. Table 3.1 summarizes the out-of-sample forecasting results for gold.

Table 3.1: AR(5) Forecasting Results - Gold

Metric	Result
MSPE AR(5)	0.00233
MSPE RW	0.00260
MSPE RW (drift)	0.00228
R^2_{Oos} vs RW (%)	10.53
R^2_{Oos} vs RW (drift) (%)	-1.88

The AR(5) model yields a lower MSPE than the random walk without drift and a positive out-of-sample R^2, indicating that past returns contain some predictive information for gold prices. This result is consistent with existing literature documenting short-term autocorrelation in gold returns and suggests superior predictive performance relative to the traditional random walk benchmark. Consequently, the strong form of market efficiency is rejected for gold returns.

In contrast, when the AR model is compared to the random walk with drift benchmark, it underperforms in out-of-sample forecasting, exhibiting a higher MSPE and a negative R^2_{Oos}. This result suggests that once a more realistic benchmark is considered, the predictive gains of the autoregressive model largely vanish. Accordingly, weak-form market efficiency for gold returns cannot be rejected.

Both results are statistically significant at the 1% level, as confirmed by the Clark and West (2007) test, which rejects the null hypothesis that the benchmark MSPE is less than or equal to that of the autoregressive model.

Overall, gold returns exhibit limited but non-zero time-series dependence. This finding aligns with the notion that precious metals markets may display weak momentum or mean-reversion effects, which are statistically detectable but economically small. Figure 3.1 provides a visual comparison of the cumulative MSPE of the AR(5) model relative to the random walk benchmark for gold.

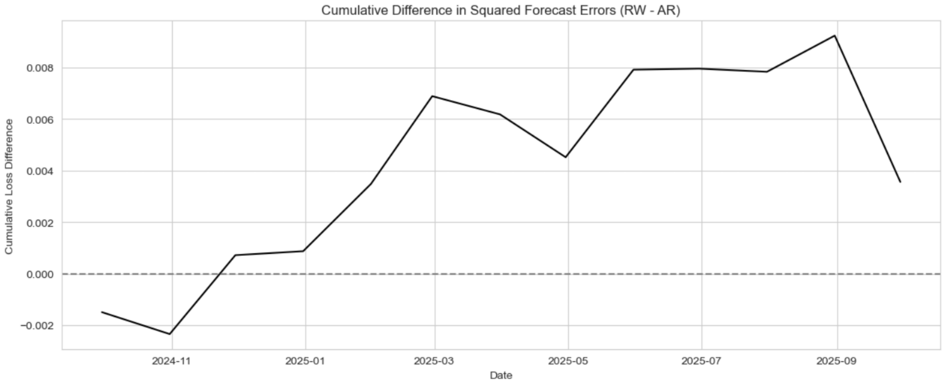

Figure 3.1: AR(5) MSPE Comparison Against Random Walk - Gold

CORN

Out-of-sample results indicate that corn returns exhibit substantial autoregressive predictability, particularly when compared to gold. The AR(5) model significantly outperforms the random walk without drift, reducing the Mean Squared Prediction Error by approximately 48%. Moreover, the autoregressive specification continues to outperform the drift-adjusted random walk benchmark, albeit with a smaller margin of about 3.8%.

Table 3.2 reports the main forecasting results for corn.

Table 3.2: AR(5) Forecasting Results - Corn

Metric	Result
MSPE AR(5)	0.00301
MSPE Random Walk	0.00585
MSPE Random Walk (Drift)	0.00313
R^2_{Oos} vs RW	48.49%
R^2_{Oos} vs RW (Drift)	3.79%

The evidence points to economically meaningful serial dependence in corn returns, leading to a rejection of market efficiency hypotheses for agricultural commodity markets. According to the Clark and West (2007) test, the null hypothesis is rejected at the 5% significance level.

The observed predictability can be rationalized by the microstructure of agricultural commodity markets. Inventory cycles, production lags, and seasonal demand patterns are consistent with persistent price dynamics. Unlike precious metals, corn prices are tightly linked to physical supply and demand constraints, which may generate gradual price adjustments and delayed information incorporation.

However, the relatively modest improvement observed when comparing the AR(5) model to the drift-adjusted benchmark suggests that incremental predictive power beyond the unconditional mean is limited. This finding highlights the importance of incorporating additional explanatory variables to better capture the fundamental drivers of agricultural commodity returns.

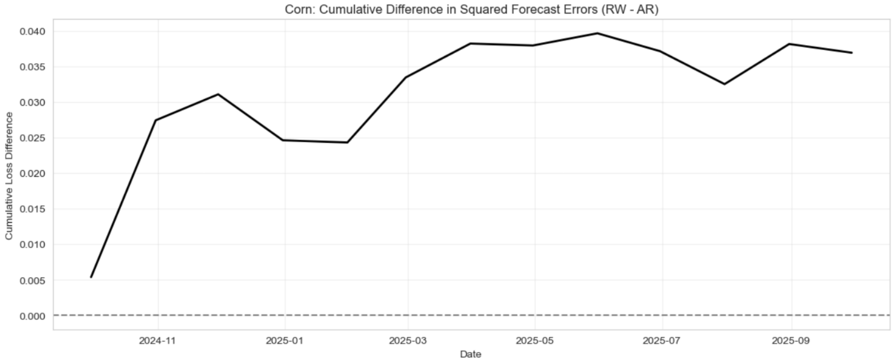

Figure 3.2: AR(5) MSPE Comparison Against Random Walk -Corn

For a visual interpretation of these results, Figure 3.2 reports the cumulative difference in squared forecast errors between the AR(5) model and the random walk benchmark. The positive and upward-sloping pattern indicates that the autoregressive model systematically improves forecasting accuracy relative to the benchmark. Therefore, despite the high volatility characterizing short-term corn returns, the AR(5) specification delivers modest but persistent gains in predictive performance over time.

NATURAL GAS

The AR(5) model for natural gas exhibits behavior similar to that observed for gold, in the sense that it substantially outperforms the random walk benchmark without drift, while underperforming the random walk with drift. Relative to the pure random walk, the autoregressive specification achieves a notably strong out-of-sample performance, with an R^2_{Oos} of approximately 56%.

By contrast, when compared to the drift-adjusted random walk, the AR(5) model delivers inferior forecasting performance, highlighting the complex dynamics characterizing natural gas markets. These markets are known to exhibit pronounced volatility clustering, regime shifts, and nonlinear responses to fundamental shocks, which are difficult to capture within a linear autoregressive framework.

Nevertheless, the superior performance relative to the traditional random walk benchmark provides evidence of strong time-series dependence in natural gas returns. This dependence may reflect the pronounced seasonal structure of energy markets, where storage constraints, weather-driven demand fluctuations, and production rigidities can generate persistent and autocorrelated price dynamics.

Table 3.3 reports the main forecasting results for natural gas.

Table 3.3: AR(5) Forecasting Results - Natural Gas

Metric	Result
MSPE AR(5)	0.02589
MSPE Random Walk	0.05941
MSPE Random Walk (Drift)	0.02473
R^2_{Oos} vs RW	56.41%
R^2_{Oos} vs RW (Drift)	-4.68%

Overall, these findings confirm that purely autoregressive dynamics provide limited forecasting power once more demanding benchmarks are considered. According to the Clark and West (2007) test, the results are consistent with weak-form market efficiency in energy markets at conventional significance levels.

These outcomes motivate the extension of the analysis in the following sections, where technical indicators and macroeconomic variables are incorporated into the forecasting framework in order to capture nonlinearities and broader sources of predictability.

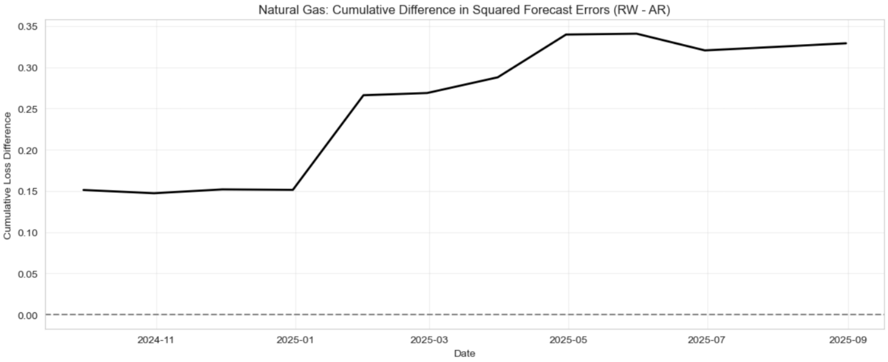

Figure 3.3: AR(5) MSPE Comparison Against Random Walk - Natural Gas

3.2 Predictive Signals and Preliminary Evidence

As a first step before continuing with the forecasting exercise, an assessment of the predictive information contained by the chosen macroeconomic and financial variables is made. To this end, an in-sample Ordinary Least Squares framework is implemented to examine the statistical and economic relationship between returns and candidate predictors. At this stage of the analysis, the purpose is not to establish genuine forecasting ability, but rather to provide evidence on whether these variables are associated with commodity price dynamics.

GOLD

Table 3.4 reports the results of the in-sample multivariate OLS regression for gold returns, where macroeconomic variables and technical indicators are jointly included as predictors. Overall, the model explains approximately 17.5% of the variation in one-period-ahead returns, with an adjusted R^2 of 10.5%, indicating a moderate but non-negligible explanatory power. Joint significance is confirmed by the F-statistic (p-value = 0.0097), suggesting statistical significance at the 1% level.

Table 3.4: In-Sample OLS Regression Results - Gold

Variable	Coefficient	Std. Error	T-stat	P-value
Constant	-0.9983**	0.465	-2.146	0.034
Avg_Mom	-4.9209	5.446	-0.904	0.368
Avg_Ma	-4.3752	3.990	-1.097	0.275
Avg_Fr	6.0640*	3.200	1.895	0.061
Avg_Oslt	8.3050**	3.598	2.308	0.023
Avg_Sr	8.3695*	4.388	1.907	0.059
T-Bill	0.3563	0.511	0.697	0.487
Long-Term Yield	-0.5234	0.834	-0.627	0.532
Exchange Rate	0.1996*	0.103	1.940	0.055
Excess Return	-0.0672	0.089	-0.756	0.451
Volatility	-0.0060	0.013	-0.466	0.642

The relevance of predictors can be discussed by distinguishing between technical and macroeconomic variables. On the technical side, the oscillator indicator is statistically significant at the 5% level and exhibits a positive coefficient, suggesting that extreme price conditions—overbought or oversold states—contribute to predicting gold returns. Filtering and support–resistance indicators are marginally significant at the 10% level. Conversely, momentum and moving average indicators do not display significant explanatory power once other variables are controlled for.

Regarding macroeconomic predictors, most variables are not statistically significant at conventional levels. However, the exchange rate shows marginal significance with a positive coefficient, consistent with the role of gold as an internationally traded asset sensitive to currency fluctuations.

Diagnostic analysis indicates that the Jarque–Bera test fails to reject residual normality, while the Durbin–Watson statistic suggests no evidence of serial correlation. Nonetheless, the large condition number signals potential multicollinearity concerns, motivating the use of regularization-based forecasting models in the subsequent sections.

CORN

The multivariate OLS model for corn exhibits slightly higher explanatory power relative to gold, with an adjusted R^2 of approximately 13%. Joint significance is confirmed by the F-statistic (p-value = 0.003), indicating that the set of predictors contains relevant information for corn returns.

Table 3.5: In-Sample OLS Regression Results -Corn

Variable	Coefficient	Std. Error	T-stat	P-value
Constant	0.583	0.837	0.696	0.488
Avg_Mom	3.4954	3.715	0.941	0.349
Avg_Ma	-0.9975	3.554	-0.281	0.779
Avg_Fr	3.9202	3.401	1.153	0.251
Avg_Oslt	10.9488*	5.647	1.939	0.055
Avg_Sr	11.3962***	3.640	3.131	0.002
T-Bill	-0.2656	0.829	-0.320	0.749
Long-Term Yield	0.172	1.462	0.118	0.907
Exchange Rate	-0.145	0.190	-0.765	0.446
Excess Return	0.1883	0.161	1.171	0.244
Volatility	0.0197	0.024	0.815	0.417

Technical indicators dominate macroeconomic predictors in explaining corn returns. In particular, the support and resistance indicator is strongly significant at the 1% level, while the oscillator is marginally significant at the 10% level. This evidence is consistent with the idea that agricultural commodity markets are influenced by speculative and technical trading behavior.

Macroeconomic variables do not exhibit explanatory power, suggesting that corn returns are primarily driven by idiosyncratic supply–demand dynamics, seasonality, and market-specific shocks rather than global macro-financial conditions.

Diagnostic tests strongly reject residual normality, with a Jarque–Bera statistic of 36.5 and excess kurtosis of 5.56, confirming the presence of fat tails and non-Gaussian behavior typical of agricultural commodity returns.

NATURAL GAS

The multivariate OLS model for natural gas yields markedly different results compared to gold and corn. The model achieves an R^2 of only 0.066 and a negative adjusted R^2, indicating that it does not improve upon a simple unconditional mean benchmark. Furthermore, the joint significance test fails to reject the null hypothesis that all coefficients are jointly equal to zero.

None of the macroeconomic or technical predictors is individually statistically significant at conventional confidence levels. This finding suggests that the selected predictors do not provide explanatory power for natural gas returns within a linear framework.

Table 3.6: In-Sample OLS - Natural Gas

Variable	Coefficient	Std. Error	T-Stat	P-Value
Const	0.6766	1.955	0.346	0.730
Avg_Mom	-3.0317	8.341	-0.363	0.717
Avg_Ma	-3.4016	4.972	-0.684	0.495
Avg_Fr	-1.8239	5.119	-0.356	0.722
Avg_Oslt	11.7888	7.530	1.565	0.120
Avg_Sr	4.1784	3.173	1.317	0.190
T-Bill	-1.6292	2.113	-0.771	0.442
Long Term Yield	1.5973	3.641	0.439	0.662
Exchange Rate	-0.1581	0.440	-0.359	0.720
Excess Return	-0.1434	0.393	-0.365	0.716
Volatility	0.0100	0.056	0.179	0.858

Importantly, this result also indicates that the model does not spuriously detect predictive patterns across all commodities, thereby increasing confidence in the validity of the significant findings obtained for gold and corn.

3.3 Forecasting Results: Expanding Window - OLS

The analysis now focuses on evaluating the genuine predictive performance of the OLS model through an out of sample framework, compared against the RW model. Through the expanding window approach, the aim is to more closely replicate real-time forecasting conditions faced by market participants. This methodology captures the dynamic nature of financial markets while ensuring that forecasts are generated using only information available at each point in time, thereby avoiding look-ahead bias.

GOLD

The expanding window OLS model delivers economically meaningful and statistically significant out-of-sample improvements relative to the random walk benchmark. Table 3.7 reports the main performance metrics.

Table 3.7: Expanding Window OLS Forecasting Performance - Gold

Metric	Result
MSPE OLS	0.0021
MSPE RW	0.0026
MSPE RW (drift)	0.0023
R^2_{Oos} vs RW	19.28%
R^2_{Oos} vs RW (drift)	9.81%

The Clarke and West (2007) test yields a statistic of 1.925 with a p-value of 0.039, rejecting the null hypothesis that the random walk benchmark is at least as accurate as the predictive model at the 5% significance level. Figure 3.4 illustrates the cumulative MSPE difference. These results suggest that macroeconomic and technical indicators contain genuine predictive information for gold returns, though the absolute out-of-sample R^2 remains modest, consistent with market efficiency and noisy variables.

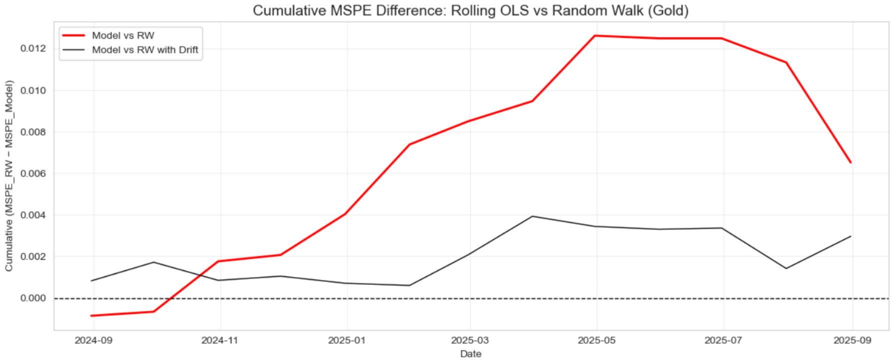

Figure 3.4: OLS MSPE Comparison Against Random Walk - Gold

CORN

The expanding window OLS model for corn shows substantial out-of-sample predictive gains relative to the random walk without drift benchmark, with $R^2_{\text{Oos}} \approx 23.2\%$. Table 3.8 summarizes the results.

Table 3.8: Expanding Window OLS Forecasting Performance - Corn

Metric	Result
MSPE OLS	0.0044
MSPE RW	0.0058
MSPE RW (drift)	0.0031
R^2_{Oos} vs RW	23.20%
R^2_{Oos} vs RW (drift)	-43.20%

The Clark–West test confirms significance of improvement relative to the non-drift random walk (test statistic = 2.33, p-value = 0.019). Negative R^2 vs RW with drift highlights that corn returns may contain a persistent unconditional mean captured well by the drift term, limiting incremental gains from multivariate predictors. Figure 3.5 illustrates the cumulative MSPE difference.

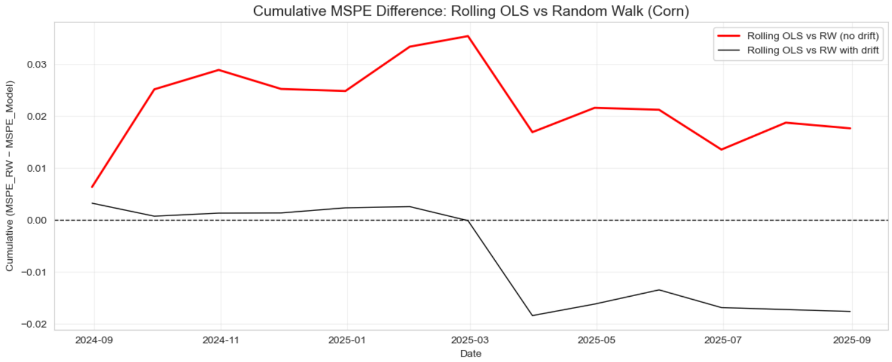

Figure 3.5: OLS MSPE Comparison Against Random Walk - Corn

NATURAL GAS

Despite weak in-sample predictors, the expanding window OLS model substantially improves out-of-sample performance relative to the random walk without drift, achieving $R^2_{\text{Oos}} \approx 49.2\%$. Table 3.9 reports the metrics.

Table 3.9: Expanding Window OLS Forecasting Performance - Natural Gas

Metric	Result
MSPE OLS	0.030
MSPE RW	0.059
MSPE RW (drift)	0.025
R^2_{Oos} vs RW	49.20%
R^2_{Oos} vs RW (drift)	-21.98%

The Clark–West test confirms significance at 1% level. However, the model underperforms relative to the random walk with drift, highlighting that macro and technical predictors provide incremental information beyond a naïve benchmark but fail to outperform a benchmark with time-varying mean. This contrast between in-sample and out-of-sample results underscores the instability of predictive relationships in energy markets and motivates the use of rolling or expanding window frameworks.

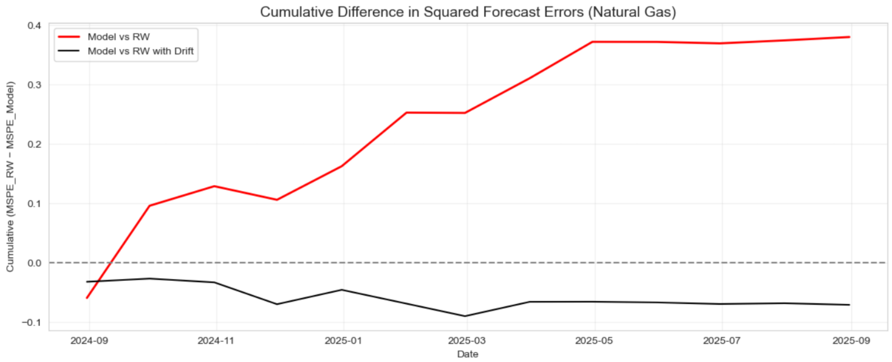

Figure 3.6: OLS MSPE Comparison Against Random Walk - Natural Gas

3.4 Forecasting Results: Expanding Window – LASSO

While the expanding-window OLS framework provides a tool to assess predictive content of macroeconomic and technical indicators, the empirical results highlight the limitations of this approach. As a matter of fact, many coefficients have proven to be weakly significant or unstable over time, and the presence of multiple correlated predictors helped raising concern over multicollinearity and overfitting issues, especially considering a forecasting context with limited time-series dimension.

The analysis is therefore extended by first implementing penalized regression techniques as the Least Absolute Shrinkage and Selection (LASSO). The goal is to improve forecasting performance with respect to OLS predictions while performing both parameter estimation and variable selection simultaneously.

Moreover, in this study, LASSO is implemented within the same expanding window forecasting framework of OLS in order to ensure fair comparison across models. Results are then measured against the random walk benchmark.

GOLD

The LASSO estimation leads to a clear improvement in out-of-sample forecasting performance for gold returns relative to the multivariate OLS model. When compared to the benchmark, the LASSO model achieves a higher r-squared, also exceeding the corresponding OLS value of 19.28%. Table 3.10 reports the main performance metrics.

Table 3.10: Expanding Window LASSO Forecasting Performance – Gold

Metric	Result
MSPE LASSO	0.0019
MSPE OLS	0.0021
MSPE RW	0.0026
MSPE RW (drift)	0.0023
R^2_{Oos} LASSO vs RW	24.38%
R^2_{Oos} LASSO vs RW (drift)	15.50%

The Clark–West test confirms significance (test statistic = 2.39, p-value = 0.017), indicating that LASSO provides meaningful gains over the benchmark. Variable selection (Figure 3.7) highlights the consistent selection of exchange rate and key technical indicators across expanding windows.

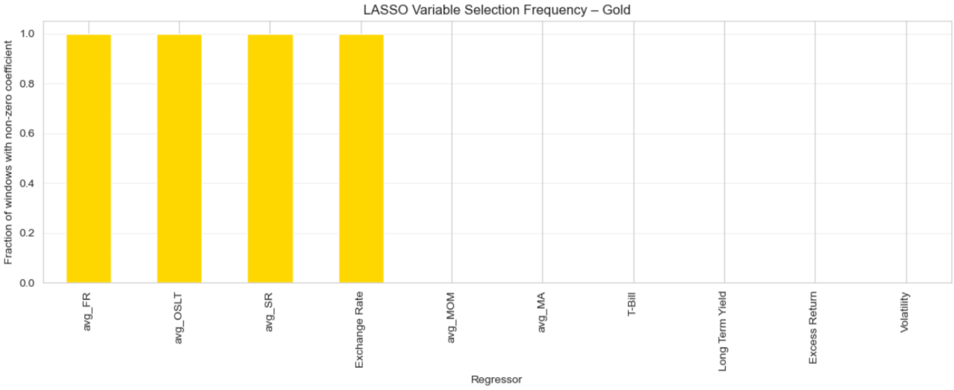

Figure 3.7: Variable selection frequency over expanding sample for gold.

CORN

The LASSO model substantially reduces forecast errors for corn relative to OLS and the random walk. Results are summarized in Table 3.11.

Table 3.11: Expanding Window LASSO Forecasting Performance – Corn

Metric	Value
MSPE LASSO	0.0038
MSPE OLS	0.0044
MSPE RW	0.0058
MSPE RW (drift)	0.0031
R^2_{Oos} LASSO vs RW	35.00%
R^2_{Oos} LASSO vs RW (drift)	-21.20%

Clark–West test confirms significance at 5%. Variable selection (Figure 3.8) shows most predictors survive LASSO regularization, though some macro variables appear only intermittently.

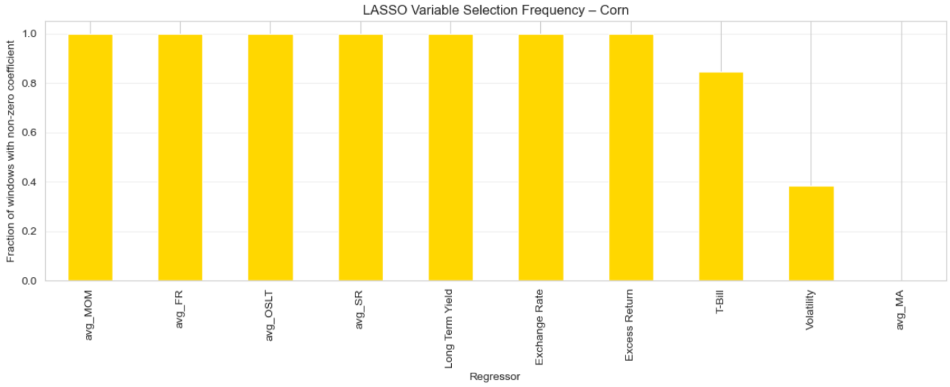

Figure 3.8: Variable selection frequency over expanding sample for corn.

NATURAL GAS

For natural gas, LASSO substantially improves out-of-sample forecasting relative to the random walk without drift, achieving a very high R^2_{Oos}. Table 3.12 presents the results.

Table 3.12: Expanding Window LASSO Forecasting Performance – Natural Gas

Metric	Value
MSPE LASSO	0.0258
MSPE OLS	0.0301
MSPE RW	0.0594
MSPE RW (drift)	0.0247
R^2_{Oos} LASSO vs RW	56.50%
R^2_{Oos} LASSO vs RW (drift)	-4.45%

Even though in-sample OLS showed no significant predictors, LASSO's regularization enhances forecast performance by reducing noise and extracting robust signals. Variable selection (Figure 3.9) shows frequent inclusion of both macroeconomic and technical indicators, reflecting the dual nature of natural gas return drivers.

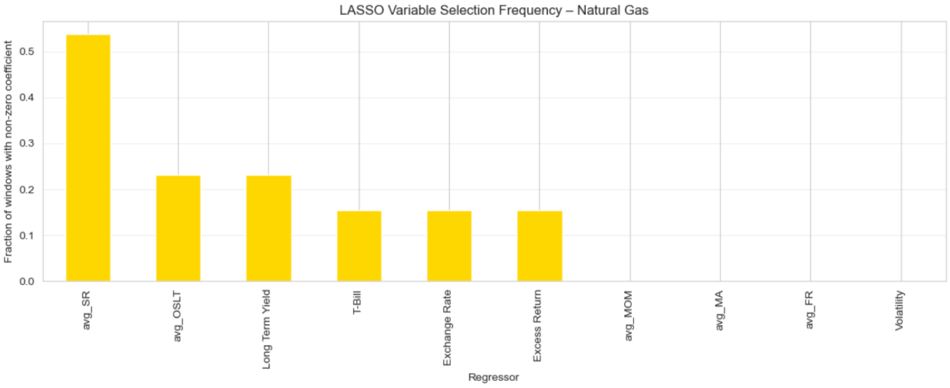

Figure 3.9: Variable selection frequency over expanding sample for natural gas.

3.5 Forecasting Results: Expanding Window – Ridge

While the LASSO results highlight the benefits of imposing sparsity by selecting only a subset of predictors when performing the forecasting exercise, they do not clarify whether the observed forecasting gains stem primarily from variable selection or more generally from regularization and coefficients' shrinkage. To disentangle these effects we next turn to Ridge regression, which penalizes coefficient magnitudes without setting them exactly to zero. The table below summarizes the key findings obtained by running the ridge regression for all three commodities.

Table 3.13: MSPE Comparison Across Forecasting Models

Commodity	Ridge	LASSO	OLS	RW	RW (drift)
Gold	0.00194	0.00196	0.00209	0.00260	0.00232
Corn	0.00420	0.00380	0.00449	0.00585	0.00313
Natural Gas	0.02690	0.02584	0.03018	0.05941	0.02474

Gold

To conclude the forecasting results, the Ridge regression results for gold confirm and strengthen evidence of predictability already observed with both OLS and LASSO models. The gains are slightly higher but comparable with the LASSO performance, confirming that gold returns benefit from regularization primarily through variance reduction rather than aggressive variable selection.

From a methodological perspective the Ridge regression appears to be well suited to the gold market where predictive relationships tend to be quite correlated rather than sparse. This is because in this multiple macro-financial variable environment most variables carry predictive information, and Ridge includes all coefficients altering solely their magnitude while mitigating multicollinearity concerns. Moreover, in light of data, gold appears overall to be the commodity for which linear predictive regressions are most effective.

Corn

For corn, Ridge regression gives a more nuanced picture. The model improves upon the random walk without drift, but still fails to perform above the drifted benchmark. Most importantly, when compared to alternative methods, an insightful result emerges: Ridge performs worst than LASSO. This result suggests that sparsity may be more important than coefficient shrinkage in the corn market. Moreover, this outcome is more consistent with the idea that corn prices are more driven by a smaller subset of economically meaningful variables, rather than a broader combination of predictors. Therefore, in this context, LASSO's ability to discard irrelevant variables is more precious than Ridge's tendency to retain them.

Natural Gas

The Ridge regression results for natural gas closely resemble, but slightly underperform, those of LASSO. This confirms that regularized predictive regressions are capable of extracting economically meaningful signals even in a market characterized by more volatility and frequent structural breaks. Once again, the weaker performance of Ridge highlights that variable selection plays a crucial role in this market, where many predictors may contribute to noise rather than signal. Overall, Ridge enhances robustness over the OLS, but the results suggest that sparse models dominate shrinkage-only approaches for natural gas forecasting, reinforcing the idea that predictability is possible but it's also fragile, and highly model dependent.

4

Discussion

Empirical results presented in the previous section reveal that commodity returns predictability is not homogeneous or stable across asset classes, but rather dependent on each commodity's economic function, market structure, and exposure to the economic regimes governing during the 2015 - 2025 period. Overall, regularized predictive regressions have shown signs of improvement in out-of-sample performance relative to traditional benchmarks, but magnitude and stability of the prediction differ greatly across gold, corn, and natural gas. The differences represent interesting break points to study mechanisms underlying each commodity specific return dynamic.

4.1 Heterogeneity in Commodity Returns' Predictability

A first striking result is the clear heterogeneity across commodity groups. Autoregressive models appear to be economically meaningful for corn and natural gas when compared to the simple random walk benchmark, yet are weaker when a drift component is included. Also for gold time series dependence is exhibited, but is more limited in magnitude. Once that technical and macroeconomic indicators are included in the analysis, predictive gains become more pronounced, particularly under regularized frameworks as LASSO and Ridge regressions.

This evidence underlines the fact that commodities should not be treated as a uniform asset class, but rather, according to specific characteristics, could be grouped in subsets that exhibit positive correlation under specific circumstances. As a matter of fact, although financialization has increased co-movement across markets, structural characteristics - such as storability, production cycles, and sensitivity to geopolitical risk - continue to generate distinct return dynamics. Moreover, the period under consideration, marked by the Covid-19 shock, the inflation surge during 2021-2023, aggressive monetary tightening, and the Russia-Ukraine war, provides a particularly informative macroeconomic landscape in which to interpret the results.

4.2 Gold: Financialized Asset and Safe Heaven Dynamics

Among the three commodities, gold is the one that exhibits the most stable and economically interpretable structure. The significance exhibited by the exchange rate and technical indicator regressors suggest that gold is influenced by both economic-financial conditions and investors' positioning. As a representative of precious metals, gold occupies a dual role: on one hand it is a real asset, used for input production processes and/or as an output luxurious good itself, and on the other a financial one, it is often adopted as a hedge against inflation and currency depreciation.

Considering the sample period 2015 - 2025, gold registered a strong upwards pressure, particularly after the Covid-19 pandemic and during the inflationary surge in 2021-2023. Moreover, these episodes coincided with heightened volatility in equity markets, and substantial fluctuations in the USD Index. All of this seems consistent with the idea of gold as a safe heaven asset, preferred by investors during uncertain times.

Focusing on the sign and significance of the exchange rate as a predictor variable in an out-of-sample setting, it appears to be consistent with a situation of strong risk aversion that dominated during the sample period. More precisely, as gold is traded in dollars, an appreciation of the USD currency would lead to a decrease in demand for gold, exerting downward pressure on its price. However, in highly uncertain economic scenarios, as is the case for the selected sample period, risk aversion might dominate, and investors prefer to extend gold position to navigate market turbulence. To support this idea of negative market sentiment, it should be noted that, during the 2022-2023 Federal Reserve's tightening phase, opportunity costs of holding gold increased due to its nature of a non-yielding asset: this should have also decreased price levels of gold, yet, once again, uncertainty and geopolitical risks can be reasonably thought to have sustained investors' demand.

On the technical indicators' side, the relevance of the oscillator and support-resistance signals further suggest that gold markets incorporate short term behavioral and positioning effects. This is consistent with an increasing participation of institutional investors and exchange traded funds in metal markets, reinforcing the idea of a more an more financialized nature for gold.

Overall, results suggest that gold is still being perceived as a safe heaven from most market participants, but is also becoming more and more speculative in nature due to positive and potentially very attractive returns that it offers.

4.3 Corn: Inventory Cycles and Gradual Information Diffusion

Results for Corn exhibit stronger autoregressive and technical predictability relative to gold. Both the AR(5) and LASSO models deliver economically meaningful improvements when compared to the random walk benchmark.

Agricultural commodities differ from precious metals as they are directly linked to biological production cycles, weather conditions, and inventory management. Production lags and seasonal patterns generate gradual adjustments in supply, which tends to produce persistent price dynamics. In accordance with the nature of this asset class, agricultural commodities tend to respond less quickly to new information, incorporating shocks more gradually, creating exploitable serial dependence and positive return opportunities by rolling futures positions when seasonality patterns kick in.

The sample period is quite informative for economic interpretation. The Covid-19 crisis disrupted global supply chains, while the subsequent reopening phase triggered strong demand recovery. Again, in 2022 the Eastern European tensions affected grain exports, amplifying volatility in agricultural markets. All of these factors not only caused price spikes (2022) or record low levels (2020), but also contributed to persistent price movements that technical trading rules and autoregressive structures were able to capture.

Interestingly, macroeconomic predictors appear not to be significant for the sample period under consideration. This result could be interpreted as the outcome of the fact that commodities remain an asset class that is mainly driven by supply-demand fundamentals, rather than broader conditions.

Lastly, on model selection, corn return dynamics appear to be best predicted though small subsets of signals, avoiding to include those who might contribute to noise generation. Variable selection is therefore more valuable than coefficient shrinkage.

4.4 Natural Gas: Volatility, Regime Shifts, and Model Instability

Evidence on Natural Gas highlights the limits of linear predictive frameworks. While autoregressive models outperform the simple random walk benchmark, in sample multivariate OLS display weak explanatory power, and negative adjusted r-squared values.

Even though regularization improves the out-of-sample performance, predictive gains remain fragile and highly sensitive to model specification.

Natural gas markets during the 2015-2025 period were characterized by strong structural shifts. Expansion of US shale production altered global supply dynamics, while the Covid-19 shock collapsed demand. The most dramatic episode occurred in 2022, when after the Russian invasion of Ukraine triggered an energy crisis in Europe, characterized by high energy prices and unprecedented volatility.

Complex dynamics as the ones presented above are difficult to capture in a linear framework. Natural gas returns, moreover, are influenced by weather conditions, storage capacity constraints, geopolitical risk, and sudden supply disruptions. The improvement is LASSO over Ridge suggests that eliminating noisy predictors is more effective than merely shrinking them, consistent with an environment in which many variables are conditionally irrelevant, depending on the regime.

Overall, findings suggest that natural gas predictability is more episodic than persistent. Linear models can extract some information, but structural breaks and nonlinearities likely require more flexible modeling approaches.

4.5 Practical Implications

As the regularization of classic multivariate models has improved predictability of commodity returns, and considering these to be notoriously difficult to predict, it can be stated that the regularization practice helps generating incremental forecasting improvements for market participants.

From an investors' perspective, allocation strategies focused on selecting relevant macroeconomic and financial predictors may help in predictability improvements, particularly for gold and corn. However, instability observed by natural gas cautions against over-reliance on linear predictive models in volatile markets. It should be noted that also portfolio managers may benefit from these results by focusing on enhancing risk management and timing strategies without resorting to black-box models.

Conclusion

The aim of this thesis has been to investigate whether commodity returns can be predicted using a combination of macroeconomic, financial, and technical indicators through the application of modern linear machine learning models. Departing from a large share of existing literature, the analysis has focused on forecasting returns rather than prices, adopting an investor oriented perspective that treats commodities as a speculative asset class, rather than pure hedging instruments.

Using monthly data for gold, corn, and natural gas over the 2015 - 2020 sample period, the study implemented a series of progressive frameworks that gradually increased the level of sophistication of forecasting techniques, with the aim to achieve more and more accurate results. Starting from autoregressive benchmarks, the analysis moved to multivariate predictive regressions estimated via Ordinary Least Squares, and then turned toward modern regularization techniques, namely LASSO and Ridge. Forecast performance was evaluated in an out of sample setting using an expanding window approach and MSPE-metrics to compare results with random walk benchmarks.

Empirical results highlight several key findings. First, commodity returns are not homogeneous across asset classes. Gold is the most stable and interpretable in terms of predictive relationship linked to macro-financial conditions and market indicators, while corn exhibits more autoregressive and technical predictability, consistent with inventory cycles and gradual information diffusion. Natural gas remains difficult to be explained by traditional indicators, leading to require more sophistication in the forecasting exercise. The key takeaway is that commodities should not be treated as an homogeneous asset class, even in an increasingly financialized market environment.

Moreover, the analysis showed that regularization played a key role in improving out-of-sample predictive performance. While OLS may provide for intuitive results and deliver against traditional benchmarks, the performance may be undermined by multicollinearity and parameter instability issues. Regularized models appear to overcome these issues effectively. LASSO has proven to be particularly valuable in settings where predictive information is sparse and/or time varying, such as agricultural and energy markets. Ridge, by contrast, performs well when predictive content is diffuse across correlated macro-financial variables, as observed in the gold market. Overall, the findings underscore the importance of dimension reduction and coefficient shrinkage in noisy financial environments already known for weak and/or unstable predictability.

Lastly, the results observed in this paper are consistent with the evolving structure of commodity markets towards a more financialized categorization. The increasing relevance of financial and technical signals supports the financialization hypothesis according to which commodities are becoming more and more a speculative asset class that reflects investor behavior and global risk conditions.

Overall, this thesis contributes to the existing literature in several ways. Relative to studies of Gargano and Timmermann (2014), the analysis provides a return-based perspective on commodity predictability and extends the set of predictors by including technical indicators, as proposed by Wang et al. (2020). However, departing from their approach of assessing model uncertainty ex-post, this study focuses on jointly including all macroeconomic and technical variables within a regularized, but single, regression framework. This is preferred for a direct comparison between OLS, Ridge, and LASSO frameworks under identical forecasting conditions. Results show that the methodological choice matters when predicting commodity returns.

Despite the discussed contributions, the analysis is subject to several limitations that, however, can be easily addressed by future research. First of all, the sample period: in spite of being economically rich, it remains relatively short for forecasting exercises, and it includes a concentration of macroeconomic events that might affect the stability of the estimated relationships. Extending the sample period can benefit in the selection of truly significant predictor variables.

A second limitation concerns the dataset: in this study only three commodities are selected and studied, therefore asset-specific features may have altered results that are instead generalized for the commodity class of belonging. Moreover, when running explanatory regressions, this paper does not adapt the set of regressor variables according to the commodity, therefore forecasting accuracy is given up in exchange for effective model comparison. In light of the first point, future research could focus on selecting either (a) a larger set of commodities, or (b) indexes that replicate the behavior of a commodity class, i.e. precious metals, agricultural, or energy. For what concerns the second observation, a forward step in forecasting accuracy could be obtained by deeply studying each of the individual commodities' characteristics and selecting a tailored regression dataset accordingly.

Lastly, the analysis has relied only on the application of linear predictive models. This has been done for the sake of result interpretation, at the cost of losing in terms of forecasting accuracy. Extending the analysis to non-linear machine learning models, such as tree based methods, could provide deeper insights into commodity return dynamics, especially for markets characterized by strong structural changes.

In conclusion, while commodity return predictability remains limited and strongly asset specific, this thesis demonstrates that combining economic intuition with more and more sophisticated machine learning techniques yields improvements in forecasting performance. Findings emphasize that predictability in commodity markets is shaped by market structure and macro-financial regimes, highlighting the relevance of rigorous economic modeling in an increasingly complex financialized global economy.

Bibliography

Bhardwaj, G., Gorton, G., & Rouwenhorst, G. (2015). Facts and Fantasies about Commodity Futures Ten Years Later (Working Paper No. 21243). *National Bureau of Economic Research.* `https://www.nber.org/papers/w21243`

Campbell, J. Y., & Thompson, S. B. (2007). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies, 21*(4), 1509–1531. `https://doi.org/10.1093/rfs/hhm055`

Clark, T. E., & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics, 138*(1), 291–311. `https://doi.org/10.1016/j.jeconom.2006.05.023`

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics, 13*(3), 253–263. `https://doi.org/10.1080/07350015.1995.10524599`

Fama, E. & Kenneth F. (1987). Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage. *Journal of Business* 60(1), 55-73. `https://doi.org/10.1086/296385`

Gargano, A., & Timmermann, A. (2014). Forecasting commodity price indexes using macroeconomic and financial predictors. *International Journal of Forecasting, 30*(3), 825–843. `https://doi.org/10.1016/j.ijforecast.2013.09.003`

Gorton, G., & Rouwenhorst, K. G. (2006). Facts and fantasies about commodity futures. *Financial Analysts Journal, 62*(2), 47–68. `https://doi.org/10.2469/faj.v62.n2.4083`

Gorton, G., Hayashi, F., & Rouwenhorst, K. G. (2013). The fundamentals of commodity futures returns. *Review of Finance, 17*(1), 35–105. `https://doi.org/10.1093/rof/rfs019`

Tang, K., & Xiong, W. (2012). Index investment and financialization of commodities. *Financial Analysts Journal, 68*(6), 54–74. `https://doi.org/10.2469/faj.v68.`

`n6.5`

Wang, Y., Liu, L., & Wu, C. (2020). Forecasting commodity prices out-of-sample: Can technical indicators help? *International Journal of Forecasting, 36(2)*, 666–683 `https://doi.org/10.1016/j.ijforecast.2019.08.004`

A

A Closer Look into Data

As explained in *Data Presentation and Analysis, 1. Commodity Futures Prices*, different commodities exhibit different time series for their price development according to the intrinsic characteristics of the asset itself. Different levels of volatility then affect returns, with gold ensuring more stable outcomes against corn and natural gas.

Gold tends to be less volatile than corn and natural gas, and this is due to the intrinsic value of this commodity that has led market participants, over time, to recognize it as a haven investment. This translates into having not only lower volatility, but also lower skewness and slimmer tails respect to other commodities such as corn and natural gas.

A graphic representation of the distributions for each commodity is provided below.

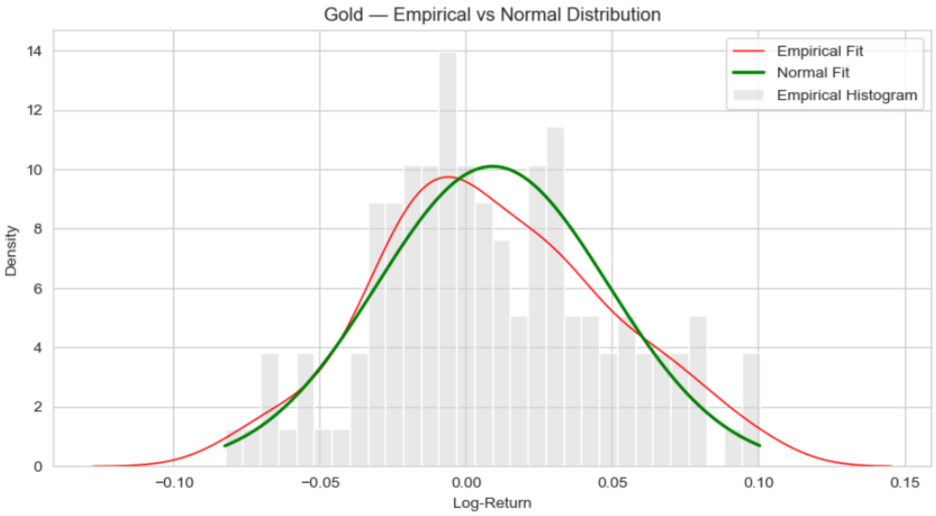

Figure A.1: Empirical vs. Theorical Distribution Comparison - Gold

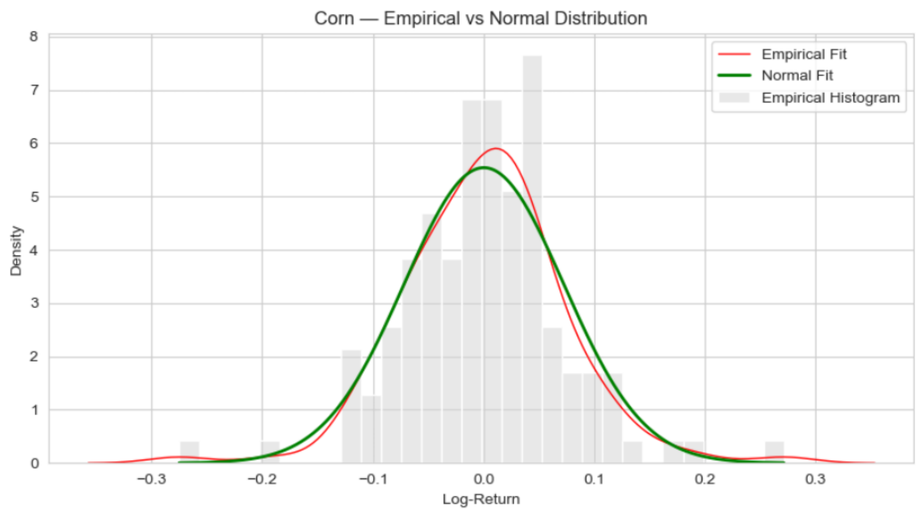

Figure A.2: Empirical vs. Theorical Distribution Comparison - Corn

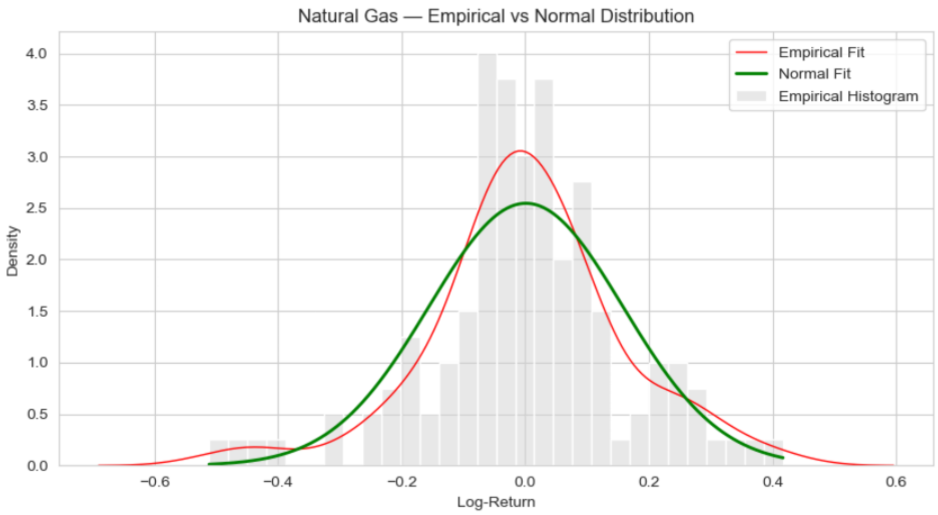

Figure A.3: Empirical vs. Theorical Distribution Comparison - Natural Gas

Moreover, to further test the hypothesis of normality the Jarque-Bera Statistic value is computed, to which higher levels correspond to lower p-values that reject the hypothesis of normality.

The results, shared below, suggest that we can confirm that only gold returns exhibit properties of normal distributions, while the p-value is at the exact break even point for natural gas, and above the threshold level that ensures normality for corn.

Table A.1: Jarque-Bera Test Results

Asset	JB Statistic	p-value
Gold	1.450755	0.484142
Natural Gas	10.577807	0.005047
Corn	25.848262	0.000002

The results are coherent with the features of each commodity class:

- Gold's distribution behaves like a normal one as the market is more stable, shocks are less likely, and volatility is contained;
- Natural gas exhibits deviations from normality due to high volatility and more frequent market shocks;
- Corn strongly rejects the normality hypothesis due to fatter tails, seasonality (typical for agricultural commodities), and high volatility.

Moreover, to further investigate properties of commodity returns, a twelve-month rolling window analysis is conducted to compute both rolling mean and standard deviation. The latter is used as a measure of local return variability, rather than an annualized volatility measure, allowing for a dynamic comparison of risk over time for each commodity. Figures 1.2, 1.4, and 1.6 plot the trends.

Results show that:

- Gold exhibits quite stable return dynamics over time, with rolling mean and standard deviation relatively low and smooth across the sample period. This behaviour reflects, once again, gold's property of being a safe haven asset for which price dynamics appear to be less affected by supply shocks or seasonal factors;
- Corn returns lie between gold an natural gas in terms of rolling standard deviation and mean behaviour. It is natural to expect rolling standard deviation higher when compared to gold due to a clear time varying behaviour driven by seasonality, harvest expectations, and sensitivity to weather conditions;
- Natural Gas displays pronounced volatility clustering, with more frequent spikes in rolling standard deviation. Due to the nature of this commodity, one can expect these volatility bursts due to exogenous shocks as weather conditions, geopolitical tensions, storage constraints, and demand fluctuations in energy markets.

On rolling means, it can be seen that these tend to fluctuate around zero, exhibiting limited predictability for all three asset classes. Overall, this rolling analysis further confirms that volatility is heteroskedastic and time-varying, with substantial differences according to the asset analyzed. These findings provide further ground for the forecasting framework adopted in the analysis, as models that assume more stable dynamics may not be well suited to capture the evolving risk structure of commodity markets.

B

Code

The complete Python code used for the empirical analysis is available at:

`https://github.com/paolarizzitano/Masters-Thesis---ML-forecasting-performance-for-commodity-returns-`

Acknowledgments

Alla mia famiglia. Grazie per avermi dato la possibilità di studiare e la libertà di costruirmi un percorso tutto mio. Quello dell'indipendenza è il valore più importante che avreste potuto trasmettermi. Vi sono grata per tutto

A mio Padre, grande esempio di curiosità intellettuale, ambizione, e senso del dovere. Grazie per essere l'esempio a cui giornalmente ambisco

Agli amici di sempre, quelli di giù. Grazie per lo spirito di casa che preservate ogni volta che ci riuniamo. Vi voglio bene

Agli amici universitari. Grazie per le giornate passate a lezione e le serate trascorse al Dome. Ricordo ancora l'entusiasmo che provavo prima di andare a dormire all'idea di rivedervi il giorno dopo. Sarò sempre grata di avervi incontrato, del gruppo che abbiamo costruito, e del bene che ci vogliamo

A Lorenzo. Ci siamo conosciuti il primissimo giorno, e già dal test di probability era chiaro avessi una marcia in più. Grazie per avermi sempre ascoltata quando condividevo con te i miei piccoli successi e, soprattutto, quando ti raccontavo dei miei fallimenti. Sei sempre stato il primo che chiamavo dopo aver finito un colloquio, in lacrime e non. Da te non mi sono mai sentita giudicata, e per questo ti ringrazierò sempre. Ti meriti tutto quello che hai raggiunto e di più! Sappi che farò sempre il tifo per te. Ti voglio bene

A Martina. Grazie per essere stata una bellissima, seppur graduale, scoperta! Come ben sai, io e te abbiamo visioni diametralmente opposte su certi aspetti della vita, ma la tua sensibilità, mista ad una personalità frizzante, non poteva che farci legare in maniera profonda. Grazie per avermi consentito di essere sempre me stessa con te presente, e sappi che, nonostante la distanza, resterò sempre la moglie da cui potrai tornare a casa a cena. Ti voglio bene

A Leonardo. Sei indubbiamente una delle persone più belle (sia dentro che fuori!) che abbia mai conosciuto. Grazie perchè sei sempre stato presente in questi anni con parole e gesti gentili nei miei confronti. Ammiro molto il tuo essere una persona spontanea e profonda, sempre pronta a dare un aiuto. Ti voglio benissimo, e sappi che potrai sempre contare su di me

A Giovanna. Lo sai, sei stata una delle primissime persone con le quali ho legato in maniera profonda. Grazie per gli infiniti giri in macchina che, tra una chiacchiera e l'altra, ci hanno portato a legare così tanto. Con il tuo spirito giocherellone e, talvolta, buffo, hai sempre saputo dare quel senso di leggerezza e vivacità alle cose. Ti voglio benissimo