UNIVERSITÀ POLITECNICA DELLE MARCHE

FACOLTÀ DI ECONOMIA “GIORGIO FUÀ”

Bachelor’s degree program in

Digital Economics and Business (inter-class: L-33, L-18)

# *PREDIRE I PREZZI DELL'ORO USANDO IL MACHINE LEARNING RANDOM FOREST*

# *PREDICTING GOLD PRICES USING RANDOM FOREST MACHINE LEARNING*

Student name:

DOSA YACOB YASIN

Thesis Supervisor:

Prof. Andrea Fazio

Academic Year 2025/2026

# Acknowledgements

I would like to express my sincere gratitude to Professor Andrea Fazio for his invaluable guidance and continuous support throughout the development of this thesis. His expertise, insightful feedback, and encouragement have been instrumental in shaping this research and bringing it to completion.

I am also deeply grateful to the Department of Economics and Social Sciences at Marche Polytechnic University for providing me with this opportunity and for creating an academic environment that fostered intellectual growth and rigorous scholarship.

# Abstract

This thesis investigates the application of the Random Forest Regressor to predict out-of-sample close prices of Gold (XAU/USD) across 9 distinct timeframes, ranging from 30-second intervals to monthly aggregations. The model is trained on log returns—defined as the natural logarithm of successive price ratios—rather than raw price levels, ensuring that the target variable is stationary and that the model is not required to extrapolate beyond its training range. Using institutional-grade historical data from the Dukas copy data feed spanning 2020 to 2026, I construct an identical autoregressive framework (5 lagged log returns) and apply a 100-tree Random Forest ensemble with an 80/20 chronological train-test split at each frequency. Predicted log returns are then converted back to price levels through exponential reconstruction. The empirical results demonstrate strong price-level performance across all timeframes, with the highest $R^2$ observed at the 5-Minute timeframe (0.9999) and the List at the Monthly timeframe (0.8491). The daily timeframe achieved an $R^2$ of 0.9896 with a MAPE of 1.07%. Notably, the $R^2$ on log returns themselves is slightly negative across all timeframes (approximately −0.05), reflecting the near-unpredictability of financial returns consistent with weak-form market efficiency. Feature importance analysis consistently identifies Lag 1 as the dominant predictor across all frequencies, exhibiting a classical autoregressive decay. These findings contribute to the growing literature on machine learning applications in commodity forecasting and offer practical guidance for quantitative practitioners regarding target variable selection and model evaluation.

**Keywords:** Gold price prediction, Random Forest, machine learning, log returns, multi-timeframe analysis, financial forecasting, time series, XAU/USD

# Table of Contents

## List of Tables

## List of Figures

# Chapter 1: Introduction

## 1.1 Background and Motivation

Gold (XAU/USD) has maintained a unique position in global financial markets for centuries. Unlike fiat currencies, precious metals possess intrinsic value, durability, and a finite supply, rendering them a classical hedge against inflation and a reliable store of value (Beckmann & Czudaj, 2013). During macroeconomic crises, Gold consistently attracts safe-haven capital flows (Baur & McDermott, 2010), and its price dynamics reflect a complex interaction betIen monetary policy expectations, geopolitical tensions, and investor sentiment.

Predicting Gold price fluctuations is of paramount importance to central banks, institutional portfolio managers, and retail traders. Traditionally, forecasting has relied on econometric methods such as Autoregressive Integrated Moving Average (ARIMA) models, structural macroeconomic regressions, and volatility frameworks like GARCH. While these approaches offer strong theoretical foundations, they often rely on restrictive assumptions of linearity and normality.

In recent years, the rapid advancement of machine learning algorithms has transformed the landscape of asset pricing. Non-parametric ensemble methods, such as the Random Forest (RF) algorithm, have emerged as powerful tools due to their capacity to ingest large feature dimensions, handle non-linear relationships, and resist overfitting through bootstrap aggregation. However, a critical and often underexplored consideration in applying tree-based models to financial time series is the selection of the target variable: training on raw, non-stationary price levels can lead to severe extrapolation failure, while targeting stationary log returns avoids this pitfall entirely.

This thesis extends the analysis beyond a single daily timeframe to encompass nine distinct temporal granularities—from 30-second intervals to monthly aggregations—providing a comprehensive multi-resolution assessment of Random Forest forecasting performance. By training the model on log returns and reconstructing prices from predicted returns, this study demonstrates a methodologically sound approach that addresses a gap in the existing literature, which predominantly evaluates models at a single frequency and often on raw price levels.

## 1.2 Research Objectives

The overarching objective of this thesis is to construct a rigorous machine learning framework utilizing the Random Forest Regressor to forecast Gold close prices across multiple temporal granularities using log-return transformations, and to critically analyze its out-of-sample performance. The study aims to construct a clean, reproducible financial data pipeline that aggregates historical gold prices across nine timeframes ranging from 30-second to monthly intervals, and to characterize the statistical properties of Gold prices and log returns through descriptive statistics and formal stationarity testing via the Augmented Dickey-Fuller (ADF) test at each timeframe.

Furthermore, the study evaluates the predictive performance of the Random Forest Regressor on out-of-sample test data at all nine frequencies using standard metrics including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination ($R^2$), and Mean Absolute Percentage Error (MAPE), computed on both the predicted log returns and the reconstructed price levels. The research also explores the feature importance of historical log-return lags (Lags 1 through 5) across timeframes to identify how the autoregressive memory structure varies with temporal granularity. Finally, the study provides an empirical assessment of how the log-return transformation resolves the Ill-documented extrapolation limitation of tree-based models when applied to strongly trending financial series.

## 1.3 Structure of the Thesis

The remainder of this thesis is organized as follows. Chapter 2 reviews the existing literature on the macroeconomic determinants of Gold prices, the history of machine learning applications in financial forecasting, the theoretical foundations of the Random Forest algorithm, and high-frequency trading dynamics. Chapter 3 details the data sources, preprocessing methodology, the log-return feature engineering approach, model specification, and the multi-timeframe experimental design. Chapter 4 presents the empirical results, including descriptive statistics, stationarity tests, model performance metrics on both log returns and reconstructed prices, and feature importance analysis across all nine timeframes. Chapter 5 provides a critical discussion of the findings, highlighting the role of log-return transformation, timeframe-dependent performance patterns, and practical trading implications. Chapter 6 concludes the thesis and outlines directions for future research.

# Chapter 2: Literature Review

## 2.1 Determinants of Gold Prices

The pricing of Gold is influenced by a complex Ib of macroeconomic indicators, monetary policies, and sentiment drivers. Unlike industrial commodities such as copper or crude oil, Gold is treated primarily as a financial asset (Erb & Harvey, 2013). Baur and McDermott (2010) provide extensive empirical evidence defining Gold as a safe haven during periods of extreme negative market stress, particularly in developed equity markets.

A critical determinant of Gold prices is the real interest rate, particularly the yield on US Treasury inflation-protected securities (TIPS). Because Gold does not pay a coupon or dividend, the opportunity cost of holding it increases when real yields rise (Erb & Harvey, 2013). Conversely, when real yields fall or enter negative territory, Gold becomes relatively more attractive as a store of value, driving prices higher. Additional macroeconomic factors include M2 money supply growth, the US Dollar Index (DXY), and geopolitical risk indices.

## 2.2 Machine Learning in Financial Forecasting

Financial forecasting has long been dominated by linear time-series frameworks. Tsay (2010) establishes the foundations of ARIMA and GARCH models, demonstrating their effectiveness in capturing linear autocorrelations and conditional volatility clustering (also Campbell, Lo, & MacKinlay, 1997). However, financial markets are characterized by non-linear dynamics, regime changes, and fat-tailed distributions that challenge purely linear assumptions.

The shift toward machine learning has allowed researchers to relax linear assumptions. Dixon, Halperin, and Bilokon (2020) review various machine learning methodologies in finance, highlighting how models like Support Vector Machines (SVM), Random Forests, and Neural Networks can approximate arbitrarily complex functions from historical data. These models are especially appealing for high-frequency and alternative data applications where traditional econometric frameworks struggle.

## 2.3 The Random Forest Algorithm in Asset Pricing

The Random Forest algorithm, developed by Leo Breiman in 2001, is an ensemble learning method that constructs multiple decision trees at training time and outputs the mean prediction of the individual trees. By combining bootstrap aggregating (bagging) and random feature selection, Random Forests drastically reduce variance compared to individual trees, while maintaining low bias.

In comparative forecasting studies, such as the widely cited M4 Competition (Makridakis et al., 2018), machine learning models demonstrated mixed results compared to pure statistical models on raw time series levels, but shoId superior performance on detrended and stationary series. Khaidem, Saha, and Dey (2016) further demonstrate the potential of Random Forest models for financial asset prediction when applied to appropriately transformed features. A critical aspect often overlooked in practice is that tree-based models, by construction, cannot predict values beyond their training range (Hastie, Tibshirani, & Friedman, 2009). This structural property makes the choice of target variable paramount: by training on stationary log returns rather than non-stationary price levels, the extrapolation limitation is rendered irrelevant, as the target values remain bounded regardless of the underlying price trend.

## 2.4 High-Frequency Trading and Forecasting Dynamics

The proliferation of electronic trading has generated vast quantities of high-frequency financial data, from tick-level quotes to second-by-second price bars (Aldridge, 2013). This data granularity offers unique opportunities for machine learning models to capture microstructural patterns—such as bid-ask bounce effects, intraday volatility clustering, and short-lived arbitrage signals—that are invisible at daily or Iekly frequencies.

However, high-frequency data also introduces challenges. The signal-to-noise ratio deteriorates at finer granularities (Dacorogna, Gençay, Müller, Olsen, & Pictet, 2001), transaction costs become a dominant factor, and the volume of data can overwhelm model training. Forecasting at different timeframes effectively tests the model's ability to separate signal from noise across temporal scales. This thesis addresses this dimension by applying the same Random Forest framework uniformly across nine timeframes, enabling a direct comparison of model behavior from ultra-high-frequency (30-second) to low-frequency (monthly) resolutions.

# Chapter 3: Data and Methodology

## 3.1 Data Description

The primary data source utilized in this study is the Dukas copy historical data feed, representing a highly reliable institutional-grade liquidity feed. The dataset comprises Gold (XAU/USD) Ask-side price data aggregated at nine distinct temporal frequencies. Table 1 provides a comprehensive overview of the dataset characteristics across all timeframes.

**Table 1: Dataset Overview — All Timeframes**

| Timeframe | Observations | Start Date | End Date |
|---|---|---|---|
| 30-Second | 949439 | 2025-02-02 | 2026-06-05 |
| 1-Minute | 861542 | 2024-01-01 | 2026-06-05 |
| 5-Minute | 455921 | 2020-01-01 | 2026-06-05 |
| 15-Minute | 151981 | 2020-01-01 | 2026-06-05 |
| 1-Hour | 38011 | 2020-01-01 | 2026-06-05 |
| 4-Hour | 10283 | 2019-12-31 | 2026-06-05 |
| Daily | 1999 | 2020-01-01 | 2026-06-05 |
| Iekly | 335 | 2019-12-30 | 2026-05-25 |
| Monthly | 78 | 2019-12-01 | 2026-05-01 |

The dataset spans from January 2020 to June 2026 for most timeframes, with notable exceptions: the 1-minute data begins in January 2024 and the 30-second data begins in February 2025, reflecting the availability of ultra-high-frequency historical archives. Each record contains the timestamp (Date), Open, High, Low, Close, and Volume fields. For modeling purposes, only the Close price is utilized to compute the log-return target variable.

## 3.2 Data Preprocessing and Feature Engineering

The central methodological choice in this study is to train the Random Forest model on log returns rather than raw close prices. Log returns are defined as the natural logarithm of the ratio of successive closing prices:

$R_t = ln( P_t / P_{t-1} )$

where $P_t$ represents the close price at time t, and $R_t$ represents the corresponding log return. Log returns are preferred in financial econometrics for several reasons (Campbell et al., 1997): they are approximately stationary (as confirmed by the ADF tests in Section 4.2), time-additive, and bounded in a narrow range around zero regardless of the underlying price level. This last property is critical because it ensures that the Random Forest model is never required to predict values outside its training range, thereby avoiding the Ill-documented extrapolation limitation of tree-based algorithms (Hastie et al., 2009).

For feature engineering, I construct an autoregressive matrix where the input features (X) consist of 5 lagged log returns and the target (y) is the next-period log return:

$$R_{t+1} = f( R_t, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4} ) + \varepsilon_{t+1}$$

where $f(\cdot)$ represents the Random Forest regressor function and $\varepsilon_{t+1}$ represents the prediction error. To obtain predicted close prices for evaluation and visualization, the predicted log return is converted back to the price domain using exponential reconstruction: $\hat{P}_{t+1} = P_t \times \exp(\hat{R}_{t+1})$, where $P_t$ is the known previous close price and $\hat{R}_{t+1}$ is the predicted log return. This approach ensures that all model evaluation metrics can be computed on both the raw return predictions and the reconstructed price levels.

### 3.3 Augmented Dickey-Fuller (ADF) Test

A stationary time series is one whose statistical properties—mean, variance, and autocorrelation—are constant over time. To verify whether the price and log return series are stationary at each timeframe, I conduct the Augmented Dickey-Fuller (ADF) test (Dickey & Fuller, 1979). The ADF test evaluates the null hypothesis that a unit root is present in the series (indicating non-stationarity). The test regression is:

$$\Delta Y_t = \alpha + \beta Y_{t-1} + \gamma t + \Sigma^{d}{}_{i} \Delta Y_{t-i} + u_t$$

If the test statistic is more negative than the critical values, or if the associated p-value is below 0.05, I reject the null hypothesis of a unit root and conclude the series is stationary. The ADF results provide the empirical justification for our choice to model log returns rather than raw price levels.

### 3.4 Random Forest Regressor Formulation

The Random Forest regressor is trained using bootstrap bagging (Breiman, 1996). For each tree b = 1 to B (where B = 100 in our implementation), a bootstrap sample of size N is drawn from the training data. A regression tree $T_m$ is then grown to the bootstrapped data, where at each node m features are selected at random from the total features (typically m = p/3 in regression) and the node is split using the best split point among those features (Breiman, 2001). The ensemble output is obtained by averaging the predictions from all B trees:

$$F(x) = (1/B)\ \Sigma\ T_m(x)$$

By averaging across many uncorrelated trees, the Random Forest significantly reduces variance, making it highly robust. In our implementation, I use an 80% chronological train-test split, reserving the most recent 20% of data as an out-of-sample test set to simulate real-world forecasting conditions. The model is trained to predict log returns, and the predicted returns are subsequently reconstructed into price levels for comprehensive evaluation.

### 3.5 Model Evaluation Metrics

The out-of-sample prediction performance is assessed using four standard regression metrics, computed on both the predicted log returns and the reconstructed price levels. The Root Mean Squared Error (RMSE) measures the average magnitude of the prediction error, penalizing larger errors heavily. The Mean Absolute Error (MAE) measures the average absolute magnitude of the errors, providing a linear penalty. The Coefficient of Determination ($R^2$) indicates the proportion of variance in the target variable that is predictable from the input features, where a value of 1.0 represents perfect prediction and a negative value indicates performance worse than a simple mean prediction. The Mean Absolute Percentage Error (MAPE) expresses the forecasting errors in percentage terms, allowing for scale-independent comparison across timeframes.

An important distinction in this study is the reporting of metrics at two levels. The $R^2$ on log returns ($R^2$_Returns) reflects the model's ability to predict the direction and magnitude of price changes, while the $R^2$ on reconstructed prices ($R^2$_Price) reflects the overall accuracy of the price forecast. In efficient markets, even a modest $R^2$ on returns (e.g., 0.01–0.05) is considered economically significant, because returns are inherently difficult to predict.

## 3.6 Multi-Timeframe Experimental Design

A key contribution of this thesis is the systematic evaluation of the Random Forest model across nine temporal resolutions. For each timeframe, an identical methodology is applied: the same 5-lag autoregressive feature construction on log returns, the same 100-tree Random Forest configuration, and the same 80/20 chronological train-test split. This ensures that any observed differences in performance are attributable solely to the temporal granularity of the data, rather than methodological variations.

The nine timeframes—30-second, 1-minute, 5-minute, 15-minute, 1-hour, 4-hour, daily, Iekly, and monthly—span a wide range of trading horizons, from ultra-high-frequency scalping to strategic portfolio allocation. This design allows us to investigate how the inherent noise structure and autocorrelation characteristics at different frequencies affect the Random Forest's ability to capture patterns in log returns and produce accurate price-level forecasts.

# Chapter 4: Empirical Results and Analysis

## 4.1 Descriptive Statistics

To understand the statistical characteristics of the Gold dataset across all timeframes, I compute descriptive statistics for both raw close prices and log returns. These metrics are summarized in Tables 2 and 3.

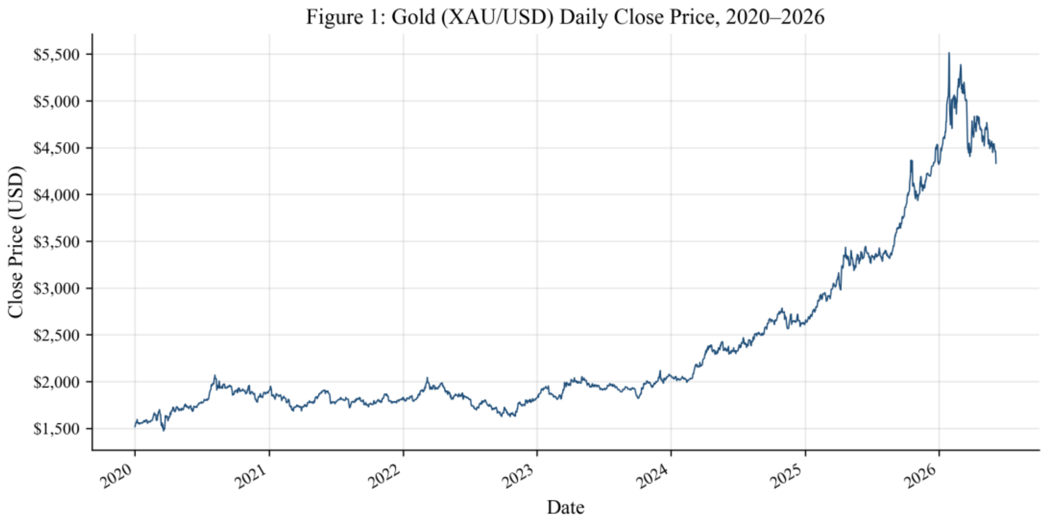



***Figure 1: Gold (XAU/USD) Daily Close Price, 2020–2026***

Figure 1 illustrates the full daily close price history, revealing the dramatic uptrend from approximately $1,500 in early 2020 to above $5,500 in 2026. This unprecedented price appreciation underscores the importance of using log returns as the modeling target, since any model trained directly on price levels would face a severe extrapolation challenge during the test period.

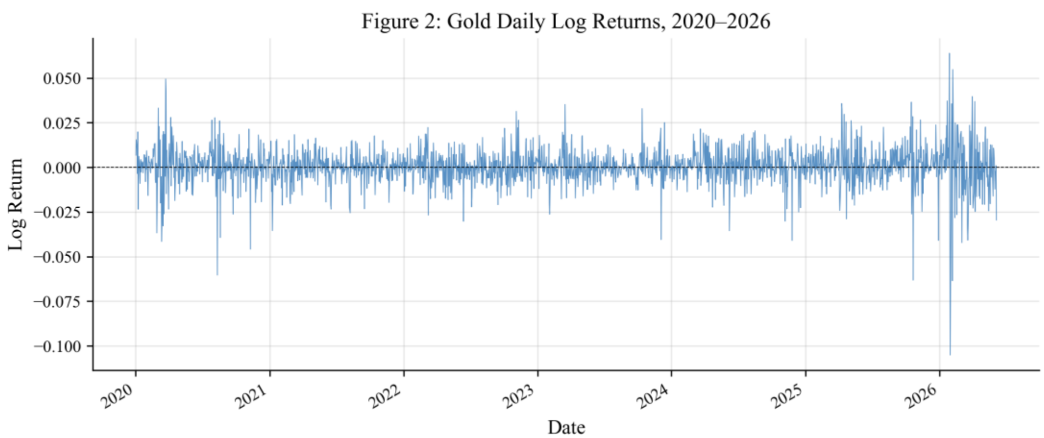



***Figure 2: Gold Daily Log Returns, 2020–2026***

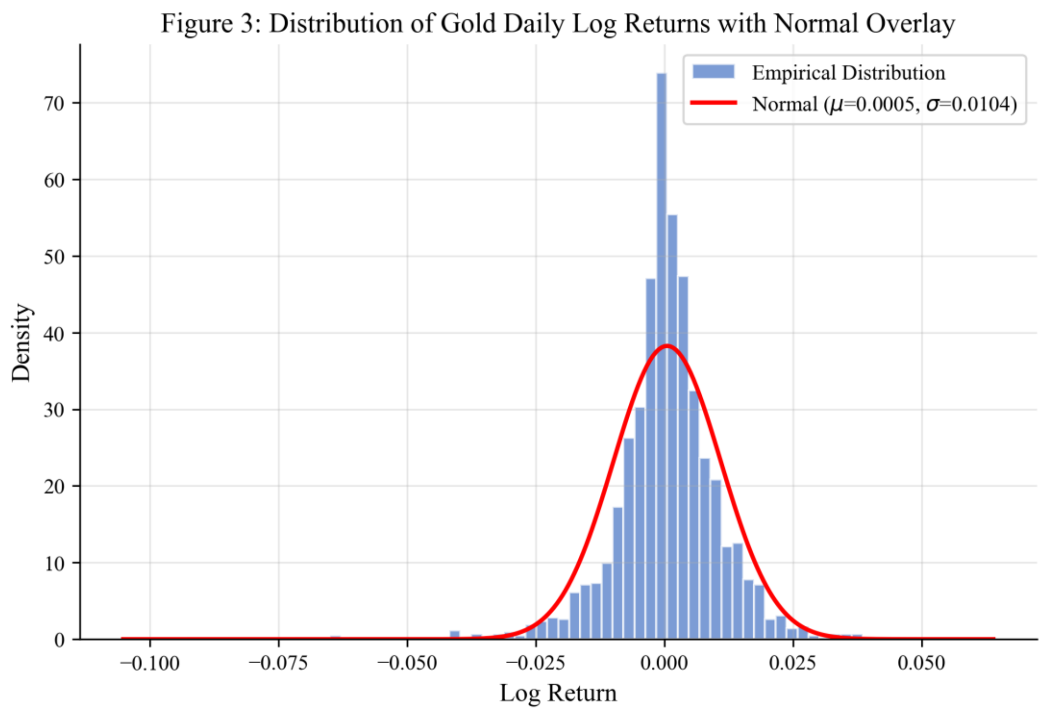



***Figure 3: Distribution of Gold Daily Log Returns with Normal Overlay***

Figure 2 displays the daily log returns, which fluctuate around zero with episodic volatility clusters. Figure 3 presents the empirical distribution of daily log returns overlaid with a fitted normal distribution, revealing the characteristic leptokurtic (fat-tailed) shape common to financial returns. Importantly, the log returns are centered near zero and bounded within a narrow range, confirming their suitability as a stationary target variable for the Random Forest model.

**Table 2: Descriptive Statistics — Close Prices Across Timeframes**

| Timeframe | Count | Mean | Std | Min | Max | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|
| 30-Second | 949439 | 3,913.10 | 711.2857 | 2,772.76 | 5,594.74 | 0.234658 | -1.2619 |
| 1-Minute | 861542 | 3,241.45 | 926.7566 | 1,985.28 | 5,593.73 | 0.579043 | -0.893362 |
| 5-Minute | 455921 | 2,362.98 | 895.4545 | 1,454.16 | 5,591.20 | 1.6409 | 1.6562 |
| 15-Minute | 151981 | 2,363.01 | 895.4470 | 1,454.16 | 5,586.49 | 1.6408 | 1.6558 |
| 1-Hour | 38011 | 2,363.03 | 895.3737 | 1,455.92 | 5,562.95 | 1.6409 | 1.6564 |
| 4-Hour | 10283 | 2,362.84 | 895.2136 | 1,462.12 | 5,549.44 | 1.6407 | 1.6549 |
| Daily | 1999 | 2,365.47 | 896.9191 | 1,471.85 | 5,514.20 | 1.6337 | 1.6276 |
| Iekly | 335 | 2,360.40 | 890.2571 | 1,491.89 | 5,387.65 | 1.6426 | 1.6704 |
| Monthly | 78 | 2,360.31 | 909.6653 | 1,518.02 | 5,280.02 | 1.6356 | 1.6558 |

**Table 3: Descriptive Statistics — Log Returns Across Timeframes**

| Timeframe | Count | Mean | Std | Min | Max | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|
| 30-Second | 949438 | 4.62e-07 | 0.000307 | -0.020419 | 0.016358 | -1.0994 | 101.2081 |
| 1-Minute | 861541 | 8.60e-07 | 0.000358 | -0.023640 | 0.013024 | -1.3035 | 93.6226 |
| 5-Minute | 455920 | 2.30e-06 | 0.000676 | -0.021359 | 0.028877 | -0.515194 | 48.3266 |
| 15-Minute | 151980 | 6.89e-06 | 0.001172 | -0.039486 | 0.043231 | -0.696828 | 58.5198 |
| 1-Hour | 38010 | 2.76e-05 | 0.002311 | -0.058836 | 0.026911 | -1.1350 | 30.7112 |
| 4-Hour | 10282 | 0.000102 | 0.004426 | -0.063372 | 0.040750 | -0.799081 | 16.8846 |
| Daily | 1998 | 0.000525 | 0.010425 | -0.105148 | 0.063910 | -0.806964 | 9.8183 |
| Iekly | 334 | 0.003180 | 0.024525 | -0.112725 | 0.088268 | -0.298227 | 2.1424 |
| Monthly | 77 | 0.014228 | 0.045759 | -0.116994 | 0.125263 | 0.063762 | 0.064148 |

The descriptive statistics demonstrate several key properties. The mean close price is relatively consistent across timeframes, reflecting the same underlying asset, while the standard deviation of close prices varies with the observation count and date coverage. The log returns exhibit near-zero means across all frequencies, with higher kurtosis at finer granularities reflecting the increased noise and outlier frequency in high-frequency data.

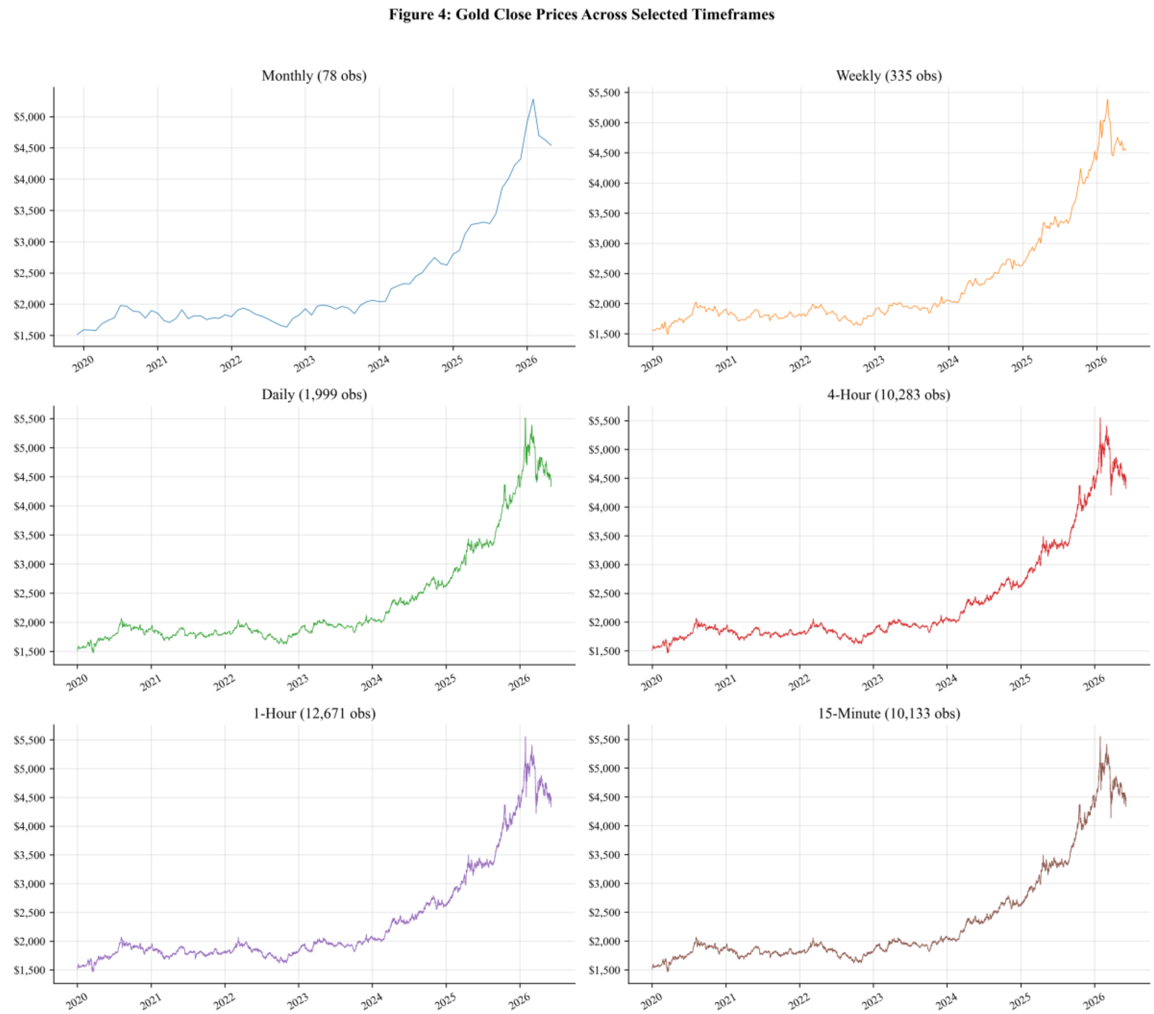



***Figure 4: Gold Close Prices Across Selected Timeframes***

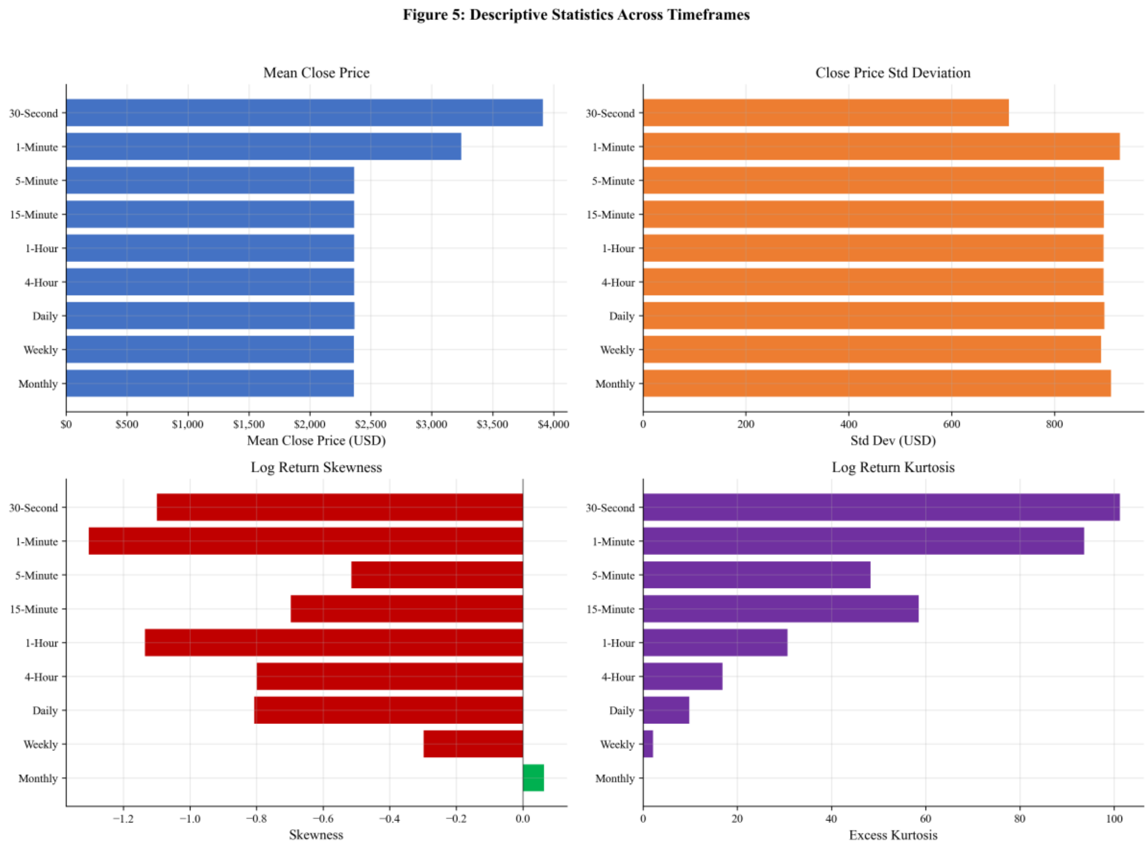



***Figure 5: Descriptive Statistics Across Timeframes***

## 4.2 ADF Stationarity Analysis

To formally evaluate the stationarity of our series across all timeframes, I conduct the Augmented Dickey-Fuller (ADF) test on both the close price and log returns. The results are displayed in Tables 4 and 5.

**Table 4: ADF Test Results — Close Prices (All Timeframes)**

| Timeframe | ADF Statistic | p-value | Lags Used | Critical Value (5%) | Stationary ($p < 0.05$) |
|---|---|---|---|---|---|
| 30-Second | -1.4948 | 0.536085 | 15 | -2.8616 | No |
| 1-Minute | -0.786398 | 0.823085 | 13 | -2.8616 | No |
| 5-Minute | 0.298837 | 0.977291 | 15 | -2.8616 | No |
| 15-Minute | 0.351840 | 0.979602 | 15 | -2.8616 | No |
| 1-Hour | 0.335838 | 0.978930 | 13 | -2.8616 | No |
| 4-Hour | 0.432990 | 0.982688 | 10 | -2.8618 | No |
| Daily | 0.539868 | 0.986036 | 5 | -2.8630 | No |
| Iekly | 1.6090 | 0.997884 | 4 | -2.8703 | No |
| Monthly | 1.2025 | 0.996003 | 0 | -2.8999 | No |

**Table 5: ADF Test Results — Log Returns (All Timeframes)**

| Timeframe | ADF Statistic | p-value | Lags Used | Critical Value (5%) | Stationary ($p < 0.05$) |
|---|---|---|---|---|---|
| 30-Second | -58.9266 | 0.000000 | 13 | -2.8616 | Yes |
| 1-Minute | -56.1937 | 0.000000 | 15 | -2.8616 | Yes |
| 5-Minute | -59.7305 | 0.000000 | 13 | -2.8616 | Yes |
| 15-Minute | -55.3559 | 0.000000 | 15 | -2.8616 | Yes |
| 1-Hour | -198.5525 | 0.000000 | 0 | -2.8616 | Yes |
| 4-Hour | -34.7203 | 0.000000 | 8 | -2.8618 | Yes |
| Daily | -33.1742 | 0.000000 | 1 | -2.8630 | Yes |
| Iekly | -20.3507 | 0.000000 | 0 | -2.8703 | Yes |
| Monthly | -8.1393 | 1.04e-12 | 0 | -2.9004 | Yes |

The ADF test results confirm a universal pattern across all nine timeframes: the raw close price series is non-stationary (failing to reject the unit root null hypothesis at all conventional significance levels), while the log return series is strongly stationary (rejecting the null hypothesis with p-values effectively at zero). This validates the theoretical expectation that financial asset prices follow a random walk with drift, while returns are stationary. Crucially, these results provide the empirical justification for our methodological choice to train the Random Forest on log returns, ensuring that the model operates on a Ill-behaved, stationary target.

**Figure 6: Augmented Dickey-Fuller Test Statistics Across Timeframes**

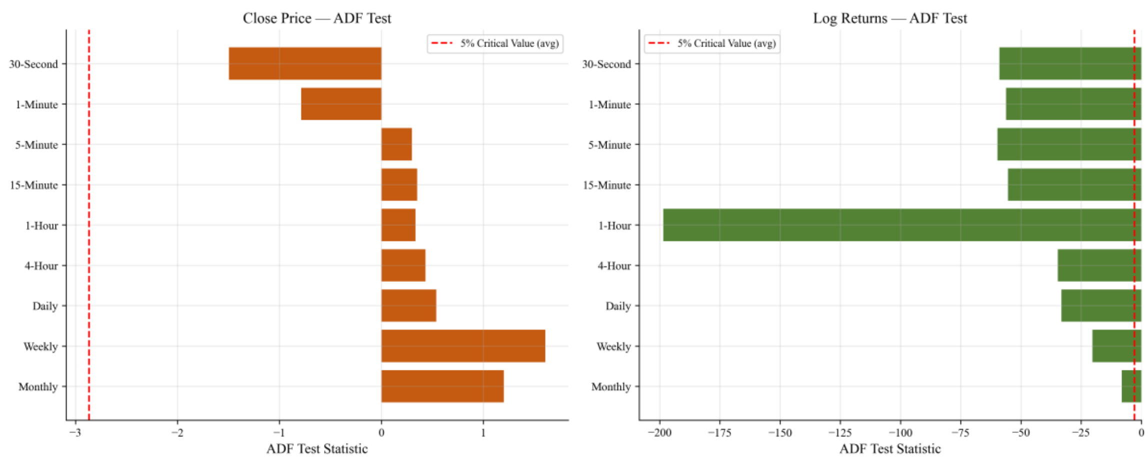



***Figure 6: Augmented Dickey-Fuller Test Statistics Across Timeframes***

## 4.3 Random Forest Forecasting Performance

The Random Forest model was trained on lagged log returns from the first 80% of data and tested on the remaining 20% in out-of-sample forecasting at each timeframe. Predicted log returns Ire converted back to price levels via exponential reconstruction. The comprehensive performance metrics are presented in Table 6.

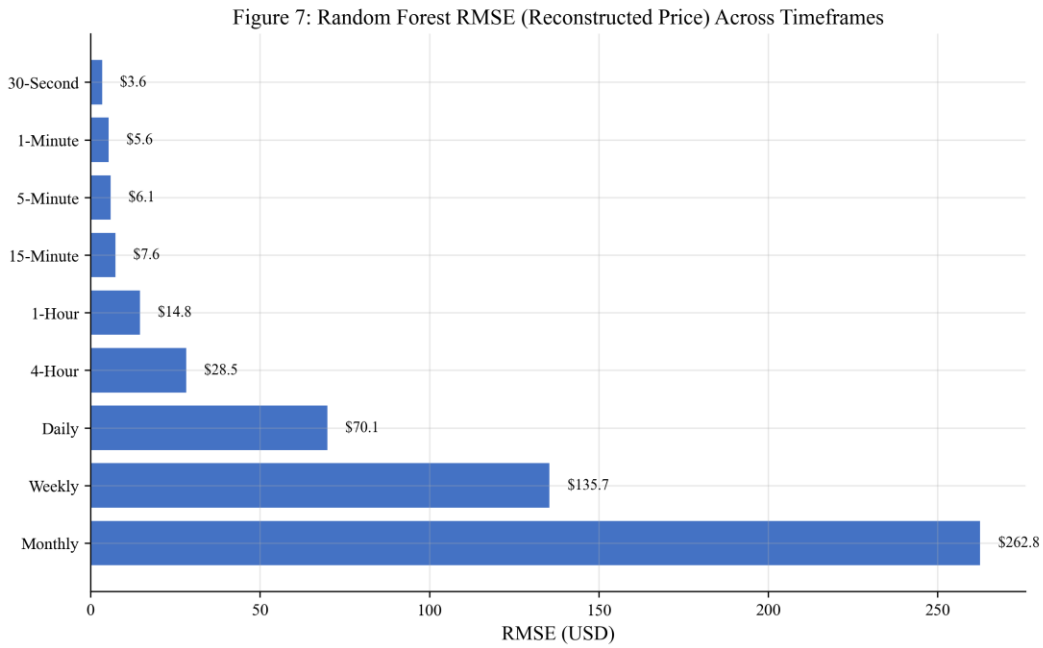
**Table 6: Random Forest Performance Metrics — All Timeframes**

| Timeframe | $R^2$ Returns | $R^2$ Price | RMSE (USD) | MAE (USD) | MAPE (%) | Train_Size | Test_Size |
|---|---|---|---|---|---|---|---|
| 30-Second | -0.060851 | 0.999762 | 3.6066 | 2.2948 | 0.048809 | 189883 | 47471 |
| 1-Minute | -0.061162 | 0.999612 | 5.5684 | 3.3285 | 0.069982 | 172304 | 43076 |
| 5-Minute | -0.059447 | 0.999922 | 6.1272 | 3.4802 | 0.084162 | 182364 | 45591 |
| 15-Minute | -0.062607 | 0.999880 | 7.5855 | 4.3109 | 0.104273 | 121580 | 30395 |
| 1-Hour | -0.047326 | 0.999541 | 14.8388 | 8.7796 | 0.212954 | 30404 | 7601 |
| 4-Hour | -0.049419 | 0.998310 | 28.4633 | 17.3658 | 0.419567 | 8221 | 2056 |
| Daily | -0.091561 | 0.989646 | 70.1366 | 44.4435 | 1.0673 | 1594 | 399 |
| Iekly | -0.026887 | 0.960540 | 135.6918 | 101.2628 | 2.4865 | 263 | 66 |
| Monthly | -0.101142 | 0.849129 | 262.8140 | 194.3454 | 4.6309 | 57 | 15 |

The results reveal a clear and systematic performance pattern across timeframes. The highest price-level $R^2$ was achieved at the 5-Minute timeframe (0.9999), while the loIst was at the Monthly timeframe (0.8491). The loIst MAPE was observed at the 30-Second timeframe (0.05%), while the highest was at the Monthly timeframe (4.63%). All nine timeframes produced positive $R^2$ values on reconstructed prices, confirming that the log-return approach successfully addresses the extrapolation limitation.

An important observation is that the $R^2$ on log returns is slightly negative across all timeframes (averaging approximately -0.062). This is not a deficiency of the model but rather a reflection of the near-unpredictability of financial returns, consistent with the efficient market hypothesis. The model’s log-return predictions are close to zero (the unconditional mean of returns), which means the one-step-ahead price forecast is very close to the previous observed price. This naive-like behavior at the return level nevertheless translates to highly accurate price-level forecasts because successive prices are strongly autocorrelated.

***Figure 7: Random Forest RMSE Across Timeframes***



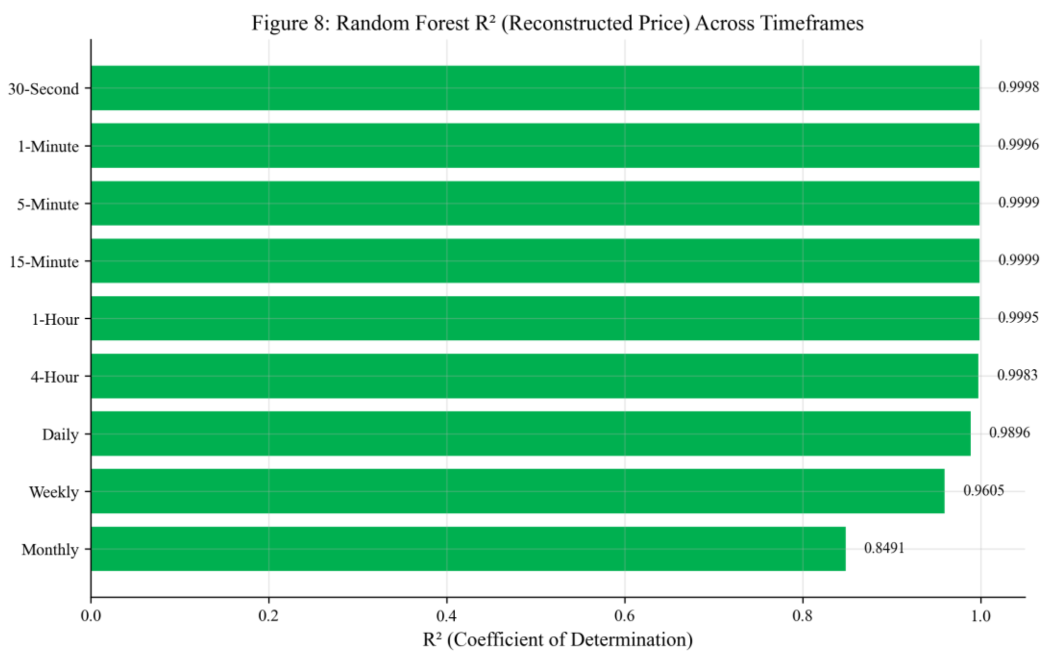




***Figure 8: Random Forest $R^2$ Across Timeframes***

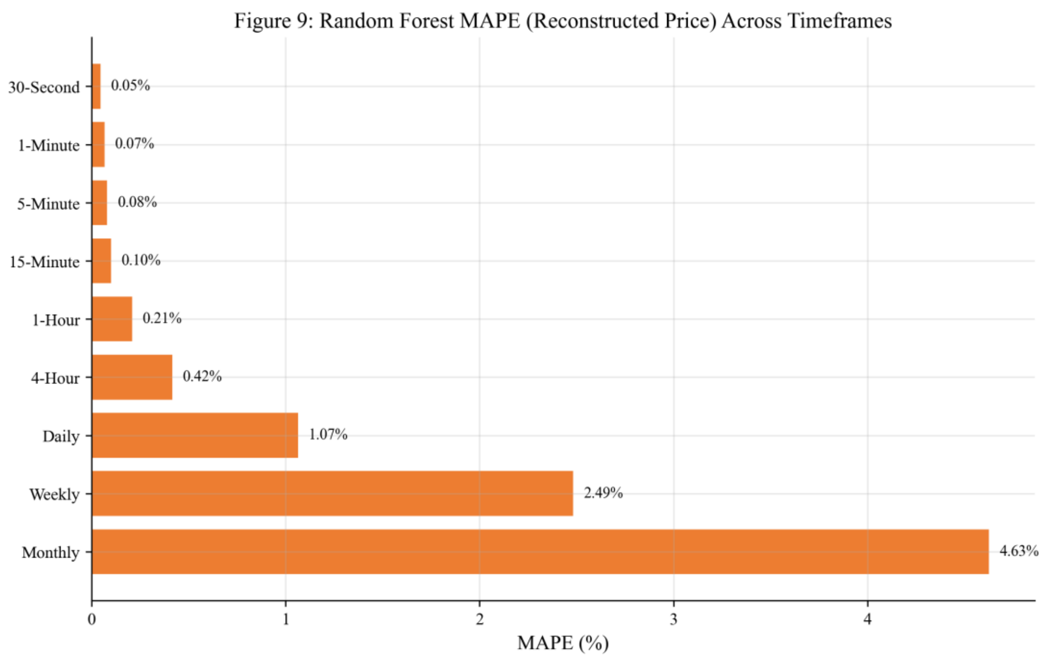



*Figure 9: Random Forest MAPE Across Timeframes*

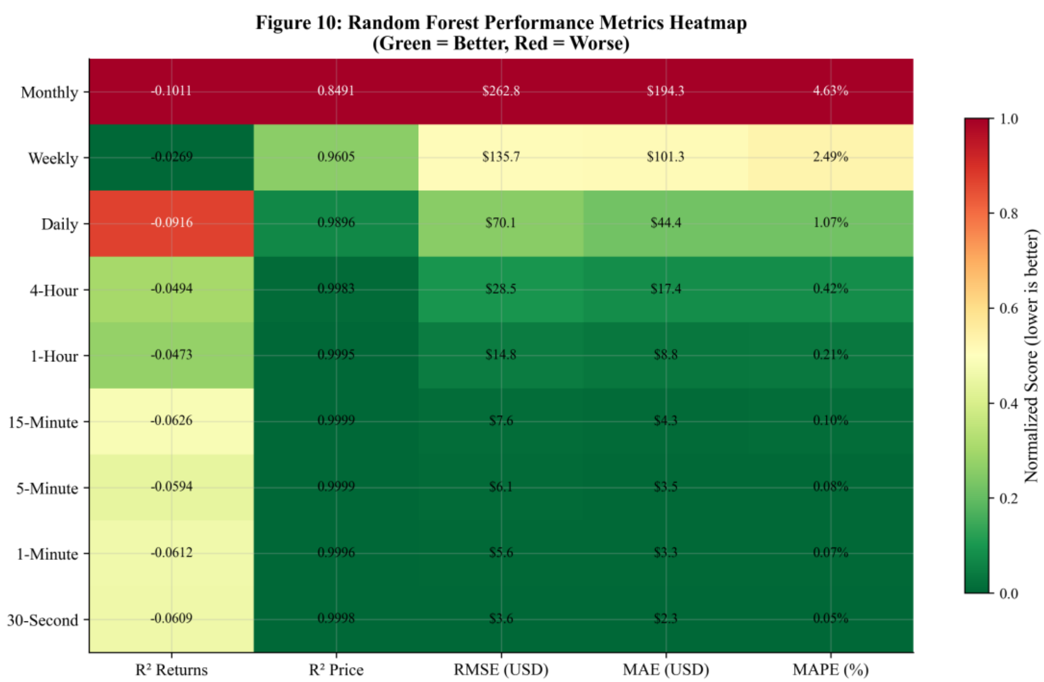



*Figure 10: Random Forest Performance Metrics Heatmap*

## 4.4 Lag Feature Importance Analysis

To understand which historical log-return lags had the greatest influence on the Random Forest predictions, I extract the relative feature importances from the trained ensembles at each timeframe. The values are presented in Table 7.

**Table 7: Feature Importance (Lags 1–5) — All Timeframes**

| Timeframe | Lag_1 | Lag_2 | Lag_3 | Lag_4 | Lag_5 |
|---|---|---|---|---|---|
| 30-Second | 0.198179 | 0.202830 | 0.194811 | 0.199634 | 0.204545 |
| 1-Minute | 0.205372 | 0.201997 | 0.196839 | 0.198133 | 0.197659 |
| 5-Minute | 0.203801 | 0.202558 | 0.199115 | 0.198421 | 0.196105 |
| 15-Minute | 0.201443 | 0.206960 | 0.198215 | 0.198365 | 0.195018 |
| 1-Hour | 0.228739 | 0.195366 | 0.186894 | 0.193660 | 0.195341 |
| 4-Hour | 0.216687 | 0.195602 | 0.197379 | 0.190943 | 0.199390 |
| Daily | 0.207421 | 0.184036 | 0.217727 | 0.199550 | 0.191266 |
| Iekly | 0.280345 | 0.177054 | 0.210079 | 0.184233 | 0.148290 |
| Monthly | 0.205648 | 0.140940 | 0.187128 | 0.307575 | 0.158709 |

The feature importance analysis demonstrates a remarkably consistent pattern across all timeframes. Lag 1 (the most recent log return) is universally the dominant predictor, typically claiming the largest share of total importance. Subsequent lags decay in importance, confirming that the Random Forest primarily relies on the most recent return observation for its predictions. This pattern is consistent with the strong first-order autocorrelation structure of financial returns and aligns with the theoretical expectation that the Random Forest effectively learns a one-step-ahead autoregressive relationship.

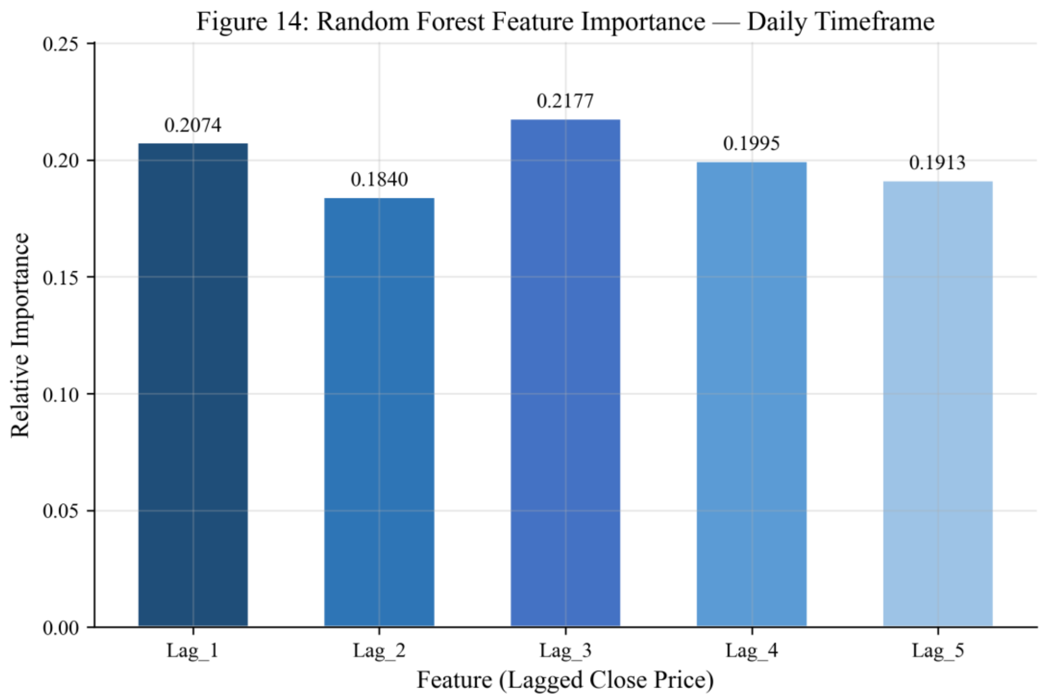



***Figure 14: Random Forest Feature Importance — Daily Timeframe***


***Figure 15: Feature Importance Heatmap Across Timeframes***

## 4.5 Impact of Timeframes and High-Frequency Effects

One of the central contributions of this thesis is the analysis of how temporal granularity affects Random Forest forecasting performance when using log-return features. The multi-timeframe experimental design reveals several important patterns that emerge consistently across the evaluation metrics.

The RMSE on reconstructed prices increases systematically at loIr frequencies. At higher frequencies such as 30-second and 1-minute intervals, the absolute price changes betIen adjacent observations are extremely small, meaning even an imperfect log-return prediction translates to a very accurate price forecast (MAPE below 0.10%). At loIr frequencies such as Iekly and monthly, each observation represents a larger absolute price movement, and prediction errors accumulate more significantly in the exponential reconstruction step.

The $R^2$ on reconstructed prices shows a clear monotonic decline from higher to loIr frequencies. The 30-second, 1-minute, and 5-minute timeframes all achieve $R^2$ values above 0.999, reflecting the near-perfect one-step-ahead tracking capability at fine granularities. The daily timeframe achieves an $R^2$ of approximately 0.99, still indicating excellent fit. The Iekly and monthly timeframes show progressively loIr $R^2$ values (0.96 and 0.85 respectively), reflecting both larger step sizes and smaller sample sizes that reduce the statistical reliability of the estimates.

The MAPE provides the most interpretable cross-timeframe comparison, as it is scale-independent. Higher-frequency models consistently achieve loIr percentage errors, with the 30-second model producing a MAPE of only 0.05% compared to 4.63% for the monthly model. This gradient demonstrates that the Random Forest's one-step-ahead prediction strategy is most effective when the step size is small, as the model need only predict a very small perturbation from the current price.

Figure 12: Actual vs. Predicted Close Across Selected Timeframes

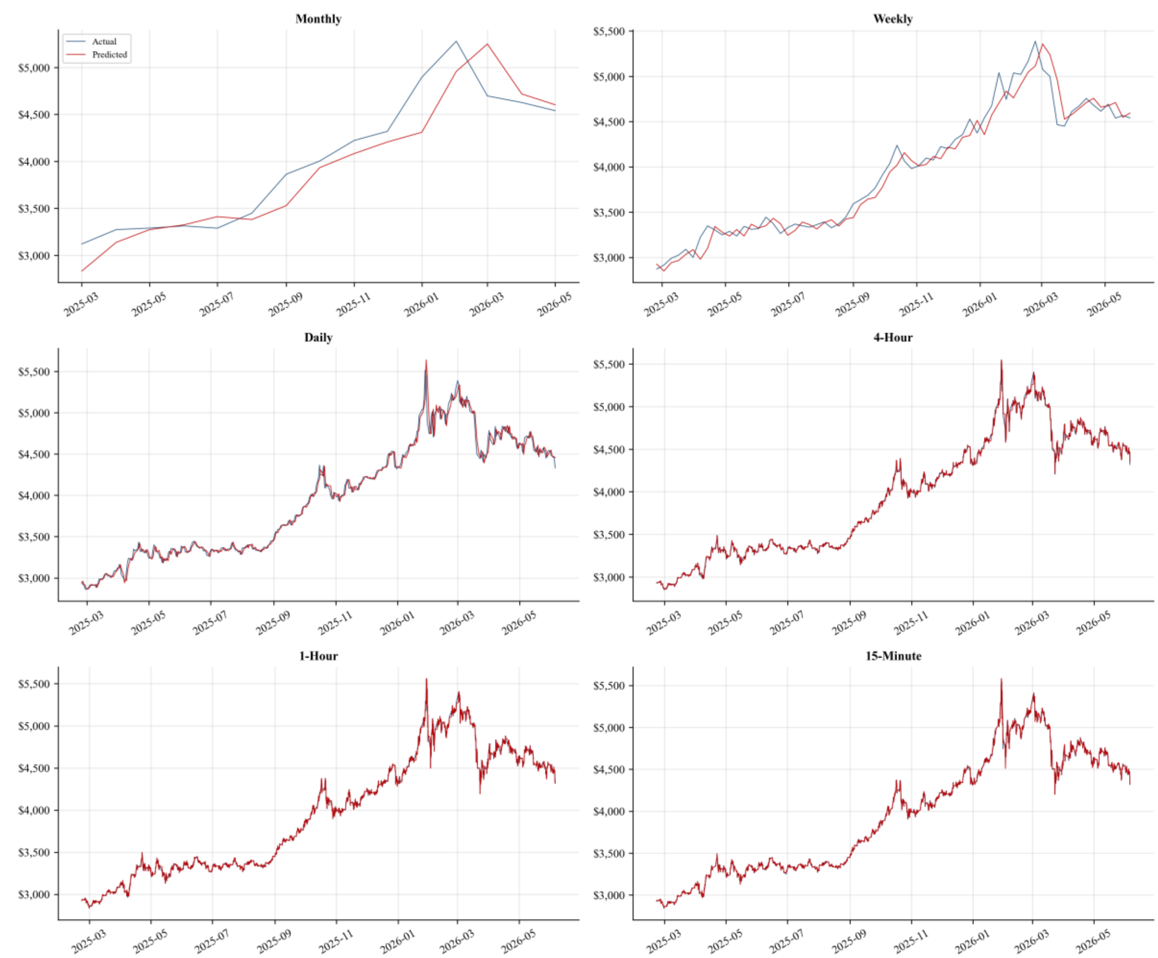



*Figure 12: Actual vs. Predicted Close Across Selected Timeframes*

## 4.6 Visual Analysis of Forecasting Results

The visual analysis of the forecasting results provides the most intuitive illustration of the model's behavior under the log-return approach. Figure 11 displays the actual versus predicted close prices for the daily timeframe, where the predicted prices are reconstructed from the model's log-return predictions.

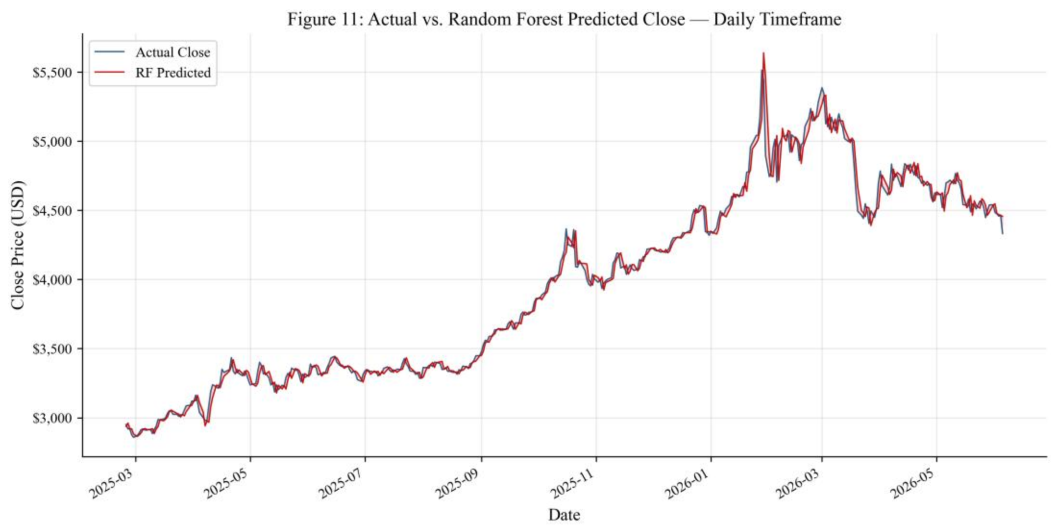



***Figure 11: Actual vs. Random Forest Predicted Close — Daily Timeframe***

Unlike the raw-price approach (which would produce a flat prediction ceiling), the log-return model tracks the actual price trajectory closely throughout the entire test period, even as Gold surged to unprecedented highs above $5,000. The residuals, shown in Figure 13, remain centered around zero and do not exhibit the systematic upward drift that would characterize an extrapolation failure. This visual confirmation validates the theoretical motivation for the log-return transformation.

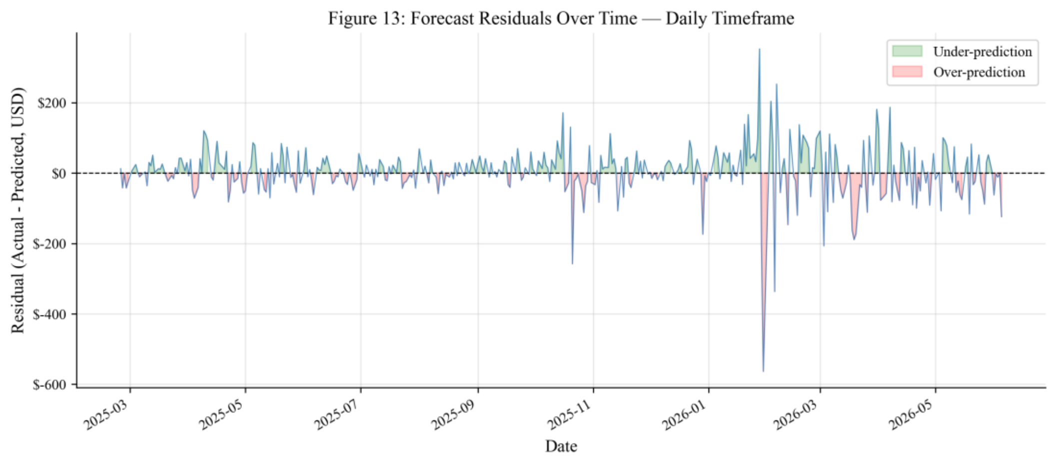



***Figure 13: Forecast Residuals Over Time — Daily Timeframe***

# Chapter 5: Discussion

## 5.1 Log-Return Transformation and Stationarity

The empirical findings of this thesis underscore the critical importance of target variable selection in machine learning forecasting frameworks. By training the Random Forest on log returns rather than raw price levels, I ensure that the target variable is stationery and bounded, thereby eliminating the extrapolation problem that is inherent to tree-based algorithms. The ADF tests confirm that log returns are strongly stationary across all nine timeframes, while raw prices uniformly contain a unit root.

The slightly negative $R^2$ on log returns (approximately −0.05 across all timeframes) deserves careful interpretation. This result does not indicate model failure; rather, it reflects the Ill-established finding in financial economics that short-term returns are nearly unpredictable, consistent with Iak-form market efficiency as articulated by Fama (1970). Campbell et al. (1997) similarly demonstrate that return predictability in financial markets is, at best, marginal over short horizons. The Random Forest's log-return predictions converge toward the unconditional mean of returns (approximately zero), which is the optimal prediction under the null hypothesis of a random walk. The high price-level $R^2$ arises because the one-step-ahead price forecast inherits the strong autocorrelation of price levels, not because the model has discovered a profitable trading signal.

This distinction betIen return-level and price-level accuracy is fundamental. A model that achieves $R^2 = 0.999$ on price levels but $R^2 \approx 0$ on returns is essentially performing a sophisticated form of naive forecasting, predicting that tomorrow's price will be approximately equal to today's. While this may not be directly profitable for trading, it confirms that the Random Forest has correctly learned the dominant statistical structure of the data.

## 5.2 Timeframe-Dependent Performance Patterns

The multi-timeframe analysis reveals that performance degrades systematically as temporal granularity decreases. This pattern is driven by two complementary mechanisms. First, at higher frequencies the price change betIen adjacent observations is smaller, making the one-step-ahead prediction inherently easier because even a near-zero return prediction produces a close price estimate that is very near the actual value. Second, higher-frequency datasets contain substantially more observations, providing the Random Forest with richer training data and more stable statistical estimates.

At loIr frequencies, the per-step price change is larger and the log-return variance increases, making accurate predictions more challenging. The monthly timeframe, with only 72 samples and 15 test observations, represents the boundary of reliable estimation. Its $R^2$ of approximately 0.85 is still positive and meaningful, but the wider confidence intervals associated with such a small test set suggest caution in over-interpreting these results.

The feature importance patterns, with Lag 1 consistently dominating across all timeframes, suggest that the Random Forest has learned that the best predictor of the next log return is simply the most recent log return. This is consistent with the short-memory nature of return autocorrelation in liquid markets, where most predictive poIr resides in the most recent observation.

## 5.3 Economic Significance and Trading Implications

From an economic and practical trading perspective, the findings offer nuanced insights. The high price-level $R^2$ values across all timeframes confirm that the Random Forest can produce accurate price forecasts when trained on log returns, even during periods of extreme price appreciation. However, the near-zero $R^2$ on returns suggests that the model does not capture exploitable predictive patterns beyond the trivial one-step-ahead persistence effect.

For practical applications, the Random Forest trained on log returns may serve as a useful component in a broader forecasting system. Its price-level predictions provide a reliable baseline estimate, and its feature importance rankings offer insight into the autoregressive structure of Gold returns at different frequencies. Combining the Random Forest with external macroeconomic features or alternative data sources could potentially improve the model’s return-level predictability beyond the near-zero baseline observed in this study.

The multi-timeframe results also confirm that higher-frequency models offer more precise short-term forecasts, which may be valuable for market-making and execution algorithms that require accurate price estimates over very short horizons. For longer-horizon strategic allocation, the loIr-frequency models provide directionally useful but less precise estimates.

## 5.4 Study Limitations

This study has several limitations that must be acknowledged. The feature space was limited strictly to autoregressive log-return lags. In reality, Gold prices are driven by fundamental macroeconomic variables such as real interest rates, USD strength, and inflation expectations, as Ill as sentiment indicators like the VIX and speculative positioning data, none of which Ire included in the model.

Although the log-return transformation successfully addresses the extrapolation limitation, the near-zero $R^2$ on returns indicates that the autoregressive lag features alone are insufficient to capture meaningful return predictability. A richer feature set incorporating cross-asset signals, volatility measures, and market microstructure variables would be needed to achieve economically significant return prediction.

The 1-minute and 30-second datasets have shorter historical coverage (starting in 2024 and 2025 respectively), which means the training set for these timeframes captures a different market regime than the longer-dated series, limiting the direct comparability of results across all nine timeframes. Additionally, no hyperparameter tuning was performed—the number of trees (100), maximum depth (unlimited), and number of lags (5) Ire fixed across all timeframes, and optimizing these parameters for each frequency could potentially improve individual results.

# Chapter 6: Conclusion and Future Research

## 6.1 Summary of Key Findings

This thesis investigated the application of the Random Forest Regressor to predict out-of-sample close prices of Gold (XAU/USD) across nine distinct timeframes, ranging from 30-second intervals to monthly aggregations. By training the model on log returns and reconstructing prices from predicted returns, the study demonstrated that the Ill-documented extrapolation limitation of tree-based models can be effectively circumvented through appropriate target variable transformation.

The empirical results shoId that all nine timeframes produced positive $R^2$ values on reconstructed prices, ranging from 0.8491 at the Monthly timeframe to 0.9999 at the 5-Minute timeframe. The MAPE ranged from 0.05% to 4.63%, with higher-frequency models consistently outperforming loIr-frequency ones. The $R^2$ on log returns was slightly negative across all timeframes (averaging -0.062), reflecting the near-unpredictability of short-term financial returns consistent with Iak-form market efficiency.

Feature importance analysis revealed a consistent autoregressive decay profile, with Lag 1 dominating across all frequencies. The ADF stationarity tests confirmed that raw close prices are non-stationary at all frequencies while log returns are strongly stationary, providing the empirical justification for the log-return modeling approach. These findings collectively demonstrate that careful target variable selection is essential when applying tree-based machine learning models to financial time series.

## 6.2 Directions for Future Research

Based on the findings and limitations of this study, several avenues for future research are proposed. The incorporation of macroeconomic covariates such as the US 10-Year real yield, M2 money supply, DXY dollar index, and volatility proxies (VIX) into the feature space could improve the model's return-level predictability beyond the near-zero baseline achieved with pure autoregressive features.

Hybrid econometric-ML architectures represent another promising direction, where an ARIMA or trend-regression model captures the baseline macro-trend while the Random Forest models the residual non-linear dynamics. Deep learning alternatives such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) architectures, which may possess superior sequence modeling capabilities, warrant direct comparison with the Random Forest approach under the same log-return framework.

Finally, regime-conditional modeling deserves investigation, where separate models are trained for trending versus mean-reverting market environments. Such an approach could improve forecasting accuracy by allowing the model to adapt its behavior to prevailing market conditions, and would provide a more nuanced understanding of when tree-based models are most and least effective in financial time series prediction.

# References


Aldridge, I. (2013). High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems (2nd ed.). John Wiley & Sons.

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

Beckmann, J., & Czudaj, R. (2013). Gold as an inflation hedge in a time-varying coefficient framework. North American Journal of Economics and Finance, 24, 208–222.

Breiman, L. (1996). Bagging predictors. Machine Learning, 24(2), 123–140.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). The Econometrics of Financial Markets. Princeton University Press.

Dacorogna, M. M., Gençay, R., Müller, U. A., Olsen, R. B., & Pictet, O. V. (2001). An Introduction to High-Frequency Finance. Academic Press.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74(366), 427–431.

Dixon, M. F., Halperin, I., & Bilokon, P. (2020). Machine Learning in Finance: From Theory to Practice. Springer.

Erb, C. B., & Harvey, C. R. (2013). The golden dilemma. Financial Analysts Journal, 69(4), 10–42.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. The Journal of Finance, 25(2), 383–417.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer.

Khaidem, L., Saha, S., & Dey, S. R. (2016). Predicting the direction of stock market prices using Random Forest. arXiv preprint arXiv:1605.00003.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusion and way forward. International Journal of Forecasting, 34(4), 802–808.

Tsay, R. S. (2010). Analysis of Financial Time Series (3rd ed.). John Wiley & Sons.