Foreign Exchange Rate Forecasting Using Statistical and Deep Learning Models

Maria Georgieva, Spencer Andersen, Kushala Rani
Department of Mathematics, Northeastern University

April 23, 2024

Abstract

This study presents an exploration of the application of several statistical and deep learning models to modeling medium-term exchange rate movements in the EUR/USD currency pair. We explore a variety of ARMA models and neural network architectures in an attempt to deliver superior out-of-sample forecasting performance than a simple random walk model or illustrate the or shed light onto the challenges to achieving this. The paper concludes with a summary and outline of current limitations and potential next steps.

1. Problem Background and Description

Exchange rates, the prices at which one country's currency can be traded in terms of another, play a pivotal role in global finance and trade. The foreign exchange (FX) market, where currencies are bought and sold, is the largest financial market globally, with a turnover of $7.5 trillion per day as of April 2022, dwarfing the daily global Gross Domestic Product (GDP) of $0.26 trillion.[1] The complexity of exchange rate dynamics presents a challenge for economists, traders, and researchers alike.

Existing forecasting models can be broadly grouped into two types: 1) structural models that reflect theorized relationships between exchange rate movements and macroeconomic variables and 2) technical analysis models that attempt to uncover patters in FX rate movements through time series analysis and other statistical and machine learning approaches. However, there is no dominant model that appears to explain FX rate movements meaningfully better than a simple random walk model, at least in the medium term. This notion was first formally introduced by researchers Richard Meese and Kenneth Rogoff in a 1983 paper and has since been widely referred to as the Meese-Rogoff Puzzle.[2] The assertions of the Meese-Rogoff Puzzle are still relevant today, and FX rates modeling is an active area of academic and professional research.

This paper aims to present the application of statistical time series analysis and deep learning approaches to modeling the medium-term movements of the EUR/USD currency pair via technical analysis. While we do not realistically expect to be able to meaningfully outperform a random walk model along out-of-sample forecasting metrics, we attempt to provide a detailed exploration of the various challenges that contribute to the complexity of the problem at hand.

2. Data

The present study utilizes daily historical closing prices of the EUR/USD currency pair from January 2000 through August 2023 (a total of 6119 observations), sourced from finance.yahoo.com. We assume an average of 252 trading days per year, which corresponds to about 24 years of daily prices in our dataset. We later split our data into training (80%) and validation (20%) subsets, leaving us with approximately 4890 observations that the different models we explore will be trained on. It is important to note that while we believe our data to be sufficiently comprehensive for our purposes, they do not reflect

---

[1] BIS Quarterly Review, Bank for International Settlements (2022)
[2] Factor Investing in Currency Markets: Does it Make Sense? Amundi Research and Macro Strategy (2019)

intra-day price fluctuations or bid-ask spreads, both of which have a potential impact on FX rate movements.

The choice to analyze the EUR/USD currency pair was driven by its status as the most traded, and therefore liquid, currency pair. Additionally, the EUR has been shown to be a momentum currency, which means that past values influence its future movements. Further, activity in the EUR/USD is not generally driven by a specific trading strategy, as is the case for currencies like the Turkish lira (TRY) or Mexican peso (MXN).[2]

3. Related Work

Past approaches to technical analysis of FX rate movements include statistical methods such as Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, as well as traditional machine learning techniques like Support Vector Machines (SVM) and Relevance Vector Machines (RVM).[3] Other techniques with applications to FX forecasting include regression analysis, spectral analysis, and state space models.[4]

In recent years, deep learning techniques such as Convolutional Neural Networks (CNNs), Deep Belief Networks (DBNs), and Long-Short Term Memory (LSTM) networks have shown promise in capturing complex patterns in exchange rate data. Hybrid ensembles, combining multiple models, have emerged as a strategy to improve forecasting accuracy. Empirical studies indicate that Recurrent Neural Networks (RNNs), particularly LSTM networks, can achieve superior performance compared to other models in capturing the nonlinear dynamics of exchange rate movements. The most commonly reported evaluation metrics are Mean Squared Error (MSE) and Mean Absolute Error (MAE).[3]

4. Statistical Models of Exchange Rates

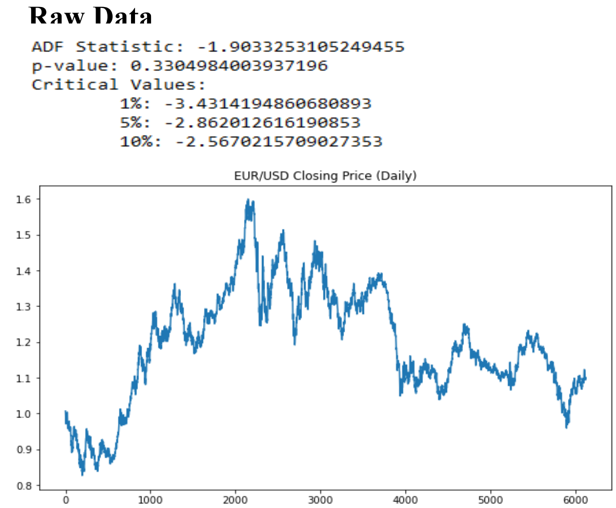
Most statistical models applied to FX forecasting are time series models – a group of models that is specifically designed to model the changes in a dependent variable that is repeatedly measured at equal time intervals. Time series data, unlike cross-sectional data, exhibit temporal dependencies where each data point is by design influenced by its preceding observations. Such dependencies introduce challenges like autocorrelation, seasonality, and heteroscedasticity of errors, which generally render the assumption of independent and identically distributed (*i.i.d.*) data invalid. Effective modeling of temporal patterns in time series analysis often requires specialized feature engineering techniques. This involves creating lagged variables, computing rolling statistics, and transforming data to handle trends and seasonality effectively.

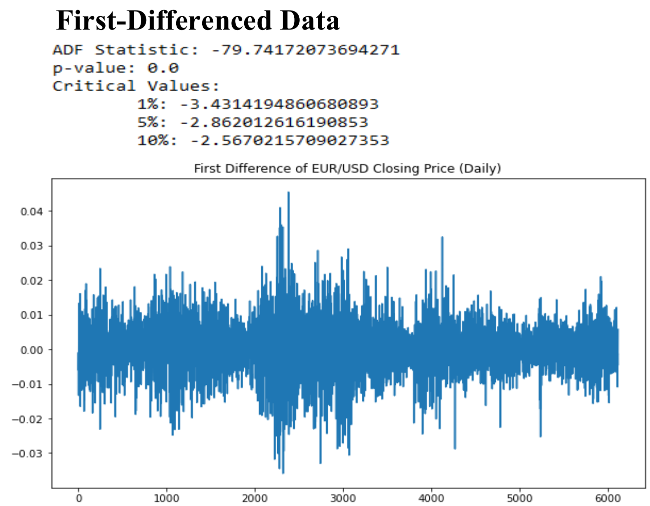
For FX time series, the choice of forecasting horizon plays a meaningful role as different models and theories hold better over a very short- or long-term horizon. Research has shown that further properties of FX time series include mean reversion, fat tails, heteroskedasticity, long memory, and nonlinearity.[2] Understanding the implications of these properties is crucial for choosing and correctly applying a model. In this paper, we focus on the SARIMAX family of models and implement several different specifications of an ARMA(p,q) model, where p and q are the autoregressive (AR) and moving average (MA) orders of the model, respectively. Given the hypothesis that FX rates follow a random walk, the last known value is chosen as the baseline model (see Appendix A for all baseline models considered). Evaluation is performed using the Mean Squared Error (MSE) metric.

---

[3] Ayitey Junior, M., Appiahene, P., Appiah, O., & Bombie, C. N. (2023). Forex market forecasting using machine learning: Systematic Literature Review and meta-analysis.
[4] Peixeiro, M. (2022). Time Series Forecasting in Python. Manning Publications Co.

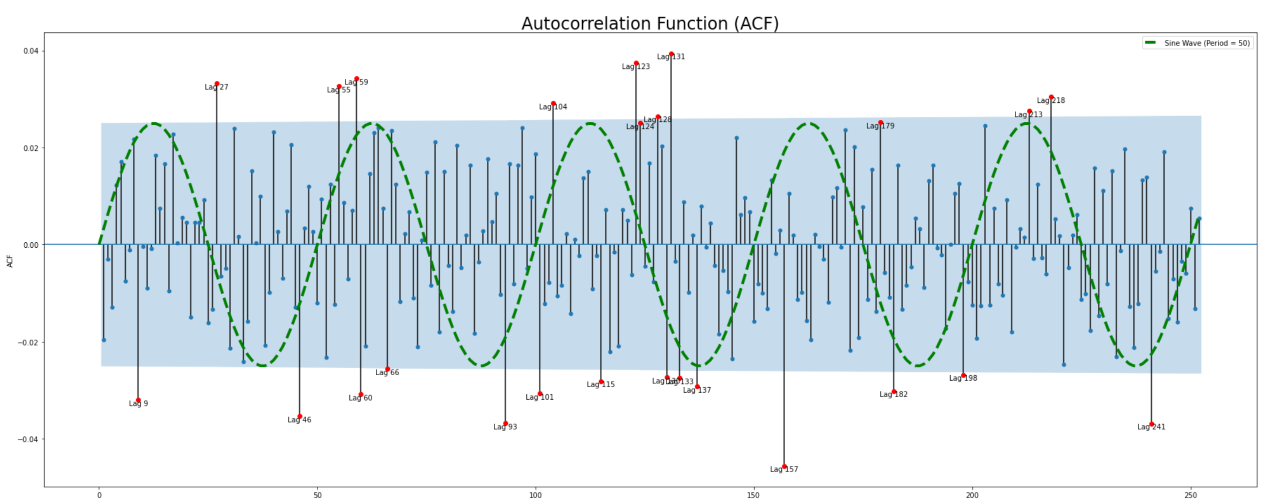
Before applying any statistical model, we explore the properties of our EUR/USD time series data. The first question to answer is whether the data series indeed appears to be a random walk and whether we can use a random walk model to describe it. In a random walk process, there is equal chance of up or down movement at any given time step. While a random walk can appear to follow upward or downward trends, to model it reliably, we need to obtain a series whose statistical properties (mean, variance) are constant over time. Such a time series is referred to as stationary. Visual inspection suggests the EUR/USD data does behave like a random walk and is not stationary. The Augmented Dickey-Fuller (ADF) test, a statistical test for stationarity, confirms this. However, first-differencing the data produces a stationary time series that we can attempt to model.





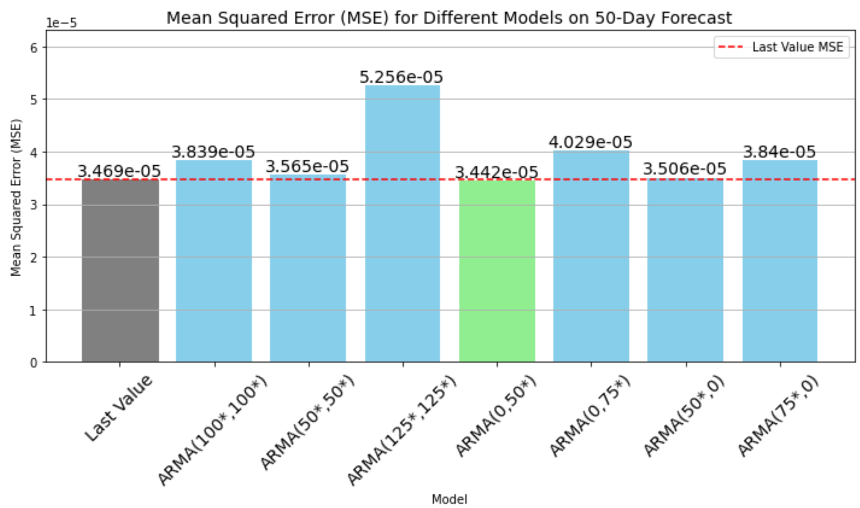
The second question we address is whether there appears to be statistically significant autocorrelation between the time series and its lagged values. To do this, we plot the autocorrelation function (ACF). Several significant lags of relatively high order (around 60 and around 130) suggest there may be some long-term temporal dependencies in the data, warranting further exploration. A decomposition of the original time series into trend, seasonal, and residual components and a periodogram further indicate mean reversion and long-term (potentially annual) periodicity (see Appendix B).

Following the generalized model selection procedure[4], we attempt to fit an ARMA(p,q) model to the data by testing various possible combinations of the orders p and q. Considering relatively low values for p and



q of up to 9 and using the Akaike Information Criterion (AIC), the best model turns out to be an ARMA(0,1) model, which is equivalent to an MA(1) model – essentially a random walk. While a residuals analysis, including qualitative evaluation with a Q-Q plot and quantitative evaluation via the Ljung-Box test show the model fits the data well and can be used for forecasting, we find evidence of significant autocorrelation of errors at higher order lags, which is consistent with the statistical property of long memory (see Appendix C).

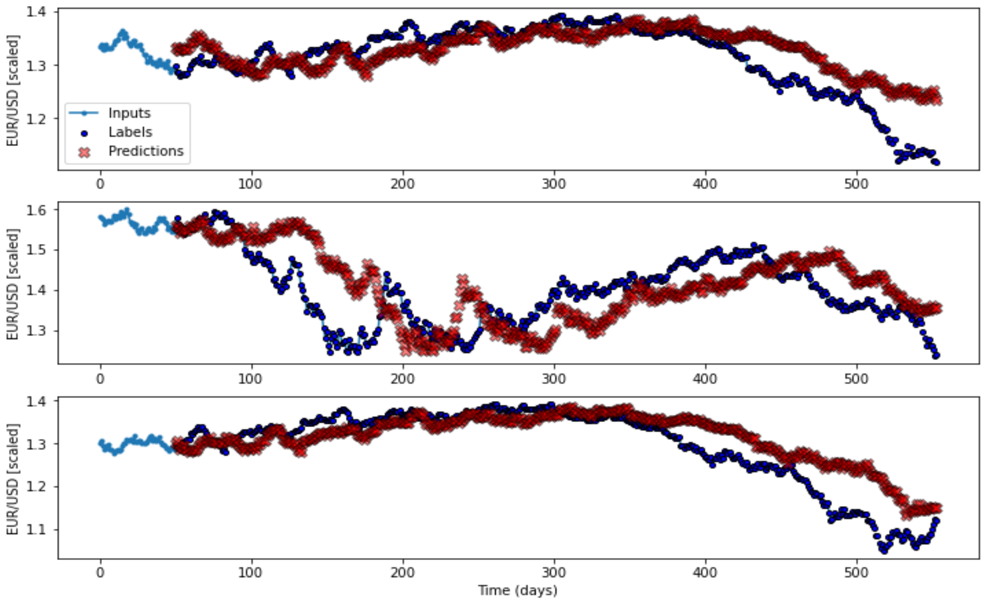
We then specify and test several ARMA(p,q) models of higher order (up to 125) that are also customized to only consider lags that have been shown as significant in the ACF plot presented above. All of these arguably more sophisticated models achieve better AIC values than the initial choice of the ARMA(0,1) model. The best performing model from this group is the ARMA(75*,0) model, where the asterisk indicates that only significant lags of order up to 75 are included in the model. Residuals analysis of this model suggests it fits the data well and does better at eliminating autocorrelation of errors at higher order lags according to the Ljung-Box test. However, the out-of-sample forecasting performance summary presented below shows that none of the models we have explored so far meaningfully outperforms the baseline.



5. Deep Learning Models of Exchange Rates

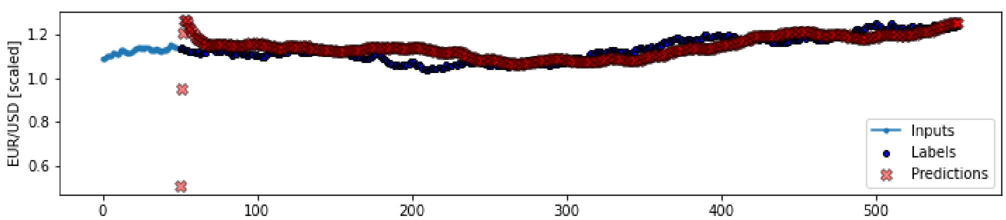
In this section of our study, we explore several different neural network architectures to attempt to capture the long memory, fat tails, and non-linearity exhibited by FX time series data. A caveat to using deep learning models in financial applications is their lack of interpretability, which makes it difficult to explain the rationale behind any investment decisions that could potentially be made based on the models.

We first implement a very simple network with just two hidden dense layers using ReLu as the activation function. To reflect our suspicion that long-term temporal dependencies are at play, we use two years (504 trading days) of prices to forecast prices for the following 50 days. Predictions made by the model based on three randomly selected sequences within our dataset are presented below.[4] The model appears to be somewhat accurate in capturing the general level of the time series but tends to diverge from the actual values whenever a directional change in trend takes place.

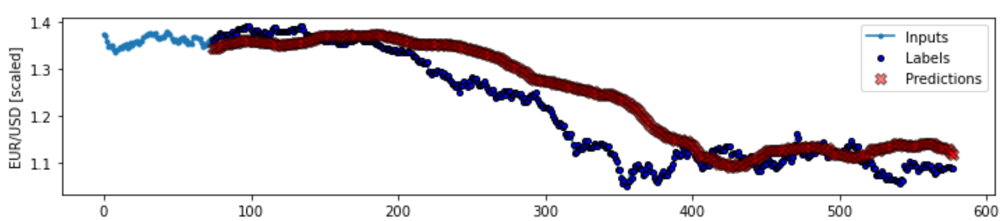


We next test several more sophisticated architectures, including:

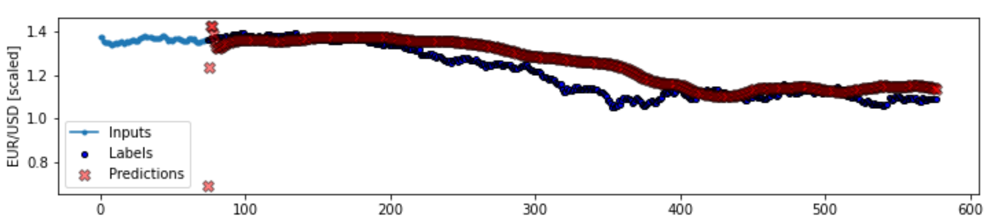
- RNN with 1 LSTM layer using 504 days of past daily values to predict the next 50 values



- CNN with a kernel width of 25 and ReLu as the activation function



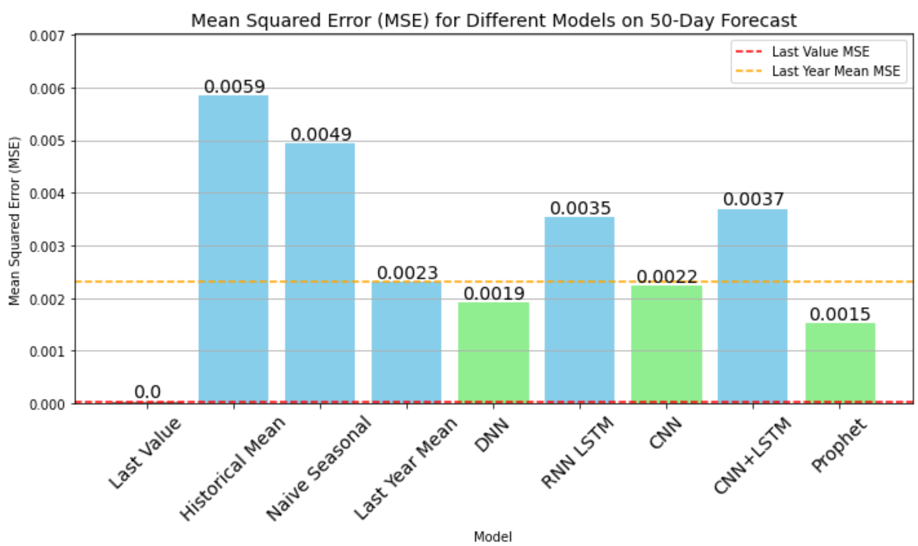
- CNN with a kernel width of 25 and 2 hidden LSTM layers



Based on visual inspection alone, none of the proposed architectures appears to fit the data particularly well. Notably, whenever we employ an LSTM layer, predictions for the first several days of the forecasting horizon are very inaccurate and likely contribute to these models' MSE disproportionately. However, because of the "black box" nature of neural networks, we do not have an interpretable understanding of why this divergence is occurring.

Finally, we applied the Prophet model – a robust, open-source forecasting tool developed by Facebook's Core Data Science team to simplify time series forecasting tasks. The model is based on an additive model and is designed to handle irregular data patterns, missing values, seasonality, trends, and holiday effects. Out of all deep learning models we implemented, Prophet delivered the best out-of-sample performance.

The below bar chart shows a summary of the out-of-sample performance of all the deep learning models we tested. They all underperformed the random walk baseline by a wide margin. Considering alternative baseline models reveals that the neural networks we implemented do outperform some naïve forecasting methods, which means that we were successful in capturing at least some of the complex patterns in the data.



6. Conclusion

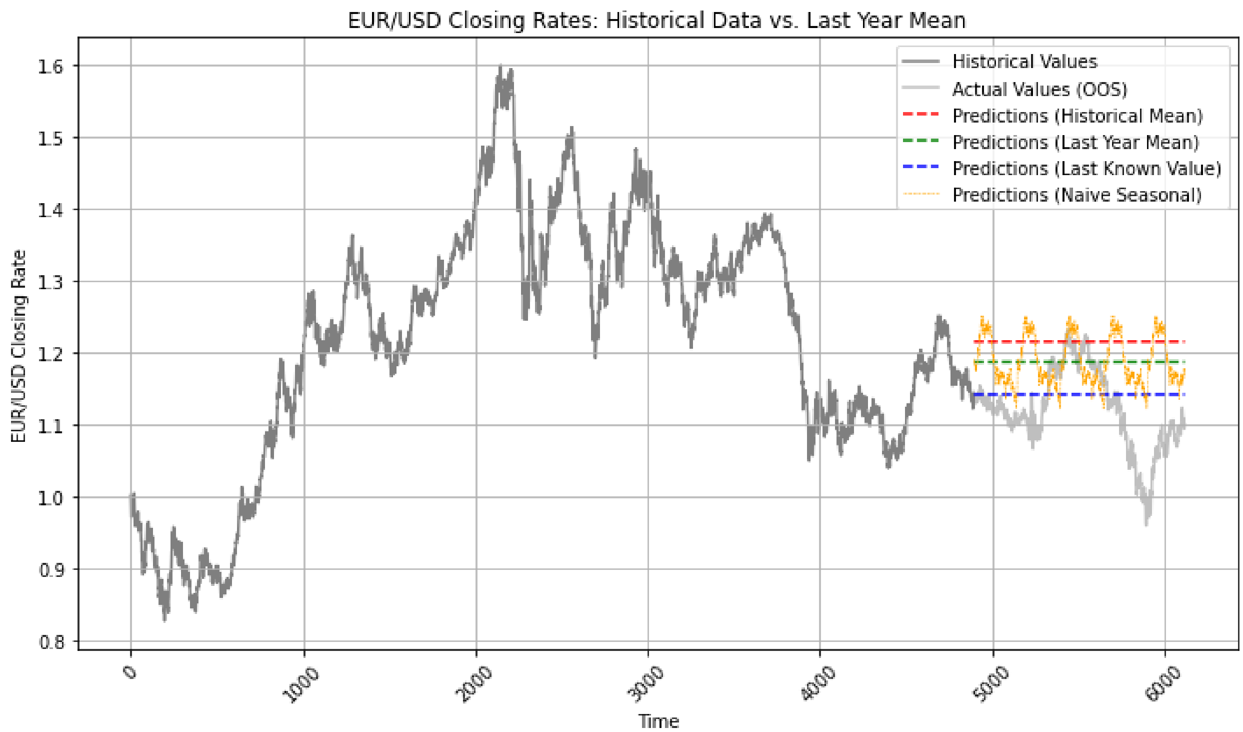
In conclusion, our analysis delves into the challenges of forecasting foreign exchange (FX) rates, particularly focusing on the EUR/USD pair. Despite employing various models, including ARMA and deep learning, we found limited success in outperforming the baseline model, which is in line with the assertions of the Meese-Rogoff Puzzle. ARMA models showed promise, emphasizing the importance of capturing autocorrelation and moving average dependencies in FX time series data. However, deep learning models yielded disappointing results, suggesting potential for improvement with more complex architectures and increased computational power.

Further research avenues include analyzing residuals with Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, leveraging machine learning for residual analysis, and incorporating exogenous variables like macroeconomic indicators. However, our study's limitations, such as the choice of data and forecasting horizon, must be acknowledged. While our analysis sheds light on FX rate forecasting challenges and opportunities, advancing our understanding and techniques is crucial for effective short and long-term forecasting, ultimately enhancing decision-making in financial markets.

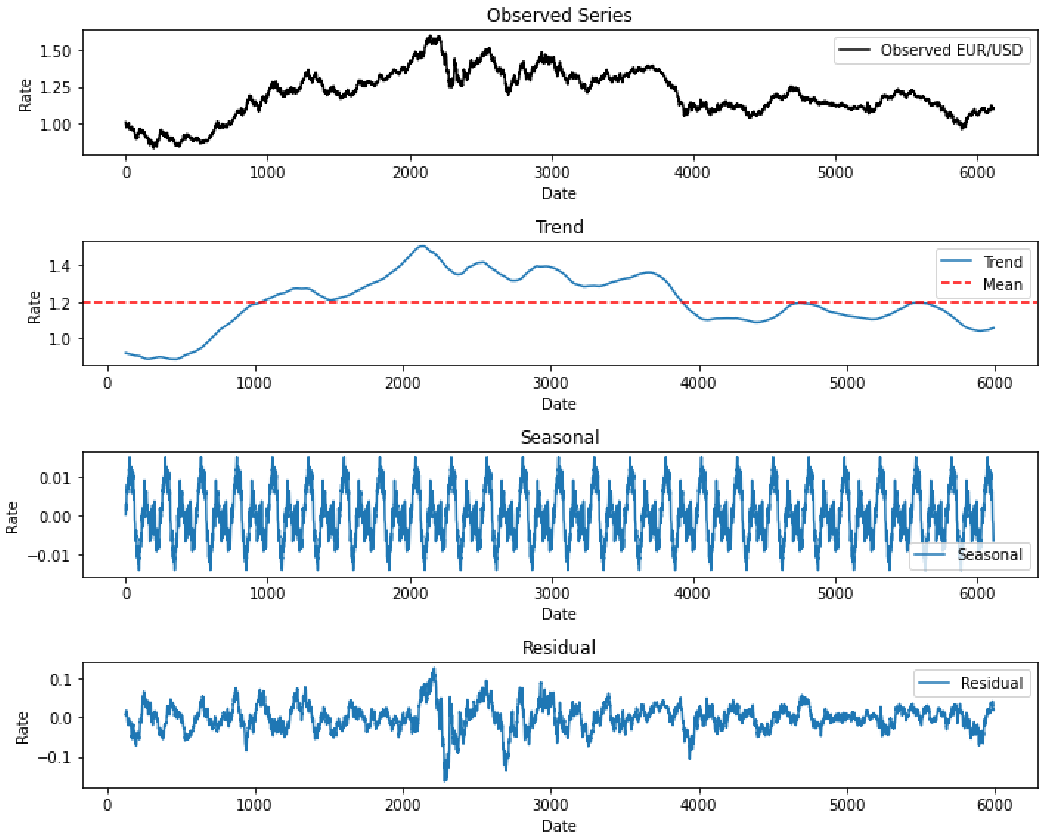
References

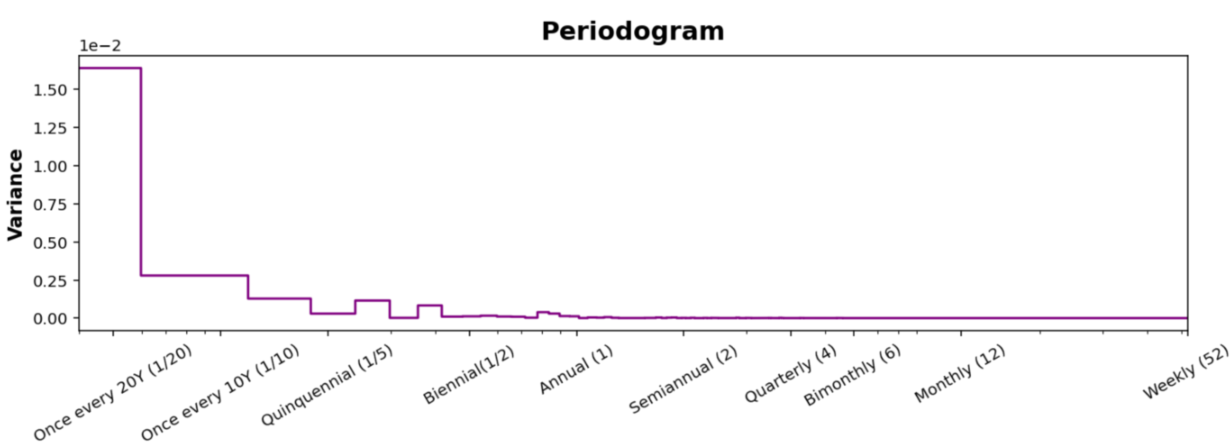
[1] BIS Quarterly Review, Bank for International Settlements (2022)
[2] Factor Investing in Currency Markets: Does it Make Sense? Amundi Research and Macro Strategy (2019)
[3] Ayitey Junior, M., Appiahene, P., Appiah, O., & Bombie, C. N. (2023). Forex market forecasting using machine learning: Systematic Literature Review and meta-analysis.
[4] Peixeiro, M. (2022). Time Series Forecasting in Python. Manning Publications Co.

Appendix A: Baseline Models

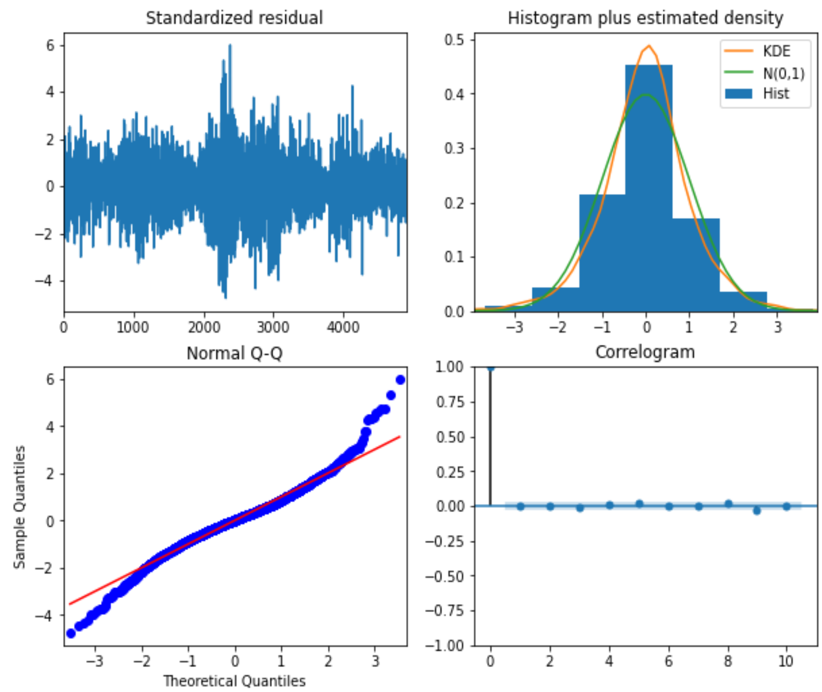


Appendix B: Time Series Decomposition

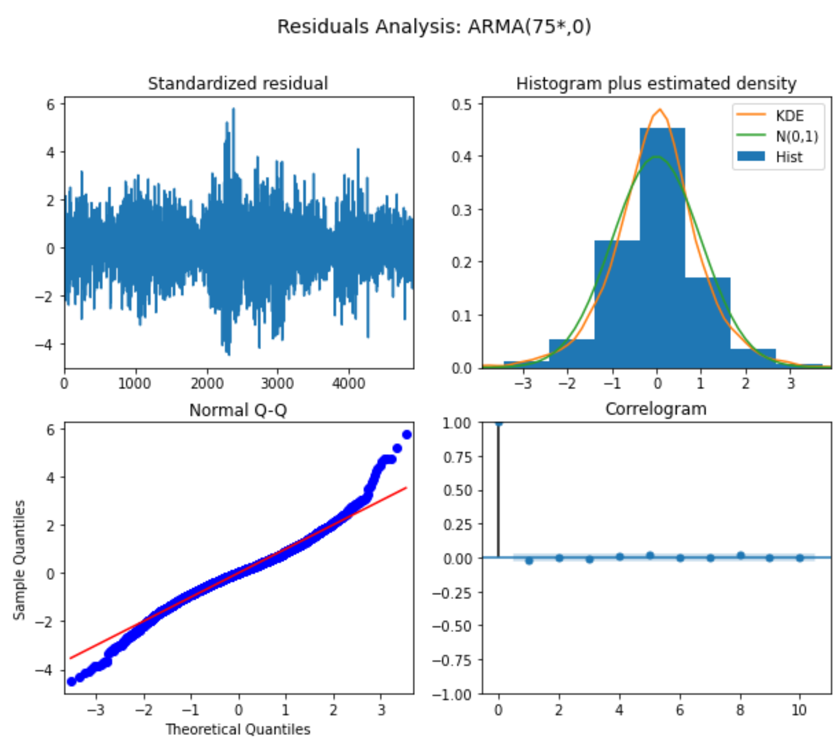




Appendix C: ARMA(0,1) Residuals Analysis



Appendix D: ARMA(75*,0) Residuals Analysis



Appendix E: ARMA(0,50*) Residuals Analysis