What can the S&P 500 returns forecast tell us about the underlying market patterns?

Final Project Paper

Authors: Indira Lakshmi Gundavarapu

DSO 522: Applied Time Series Analysis for Forecasting

Table of Contents

Abstract

Accurately forecasting stock market returns is essential for investors, analysts, and policymakers. This study aims to develop predictive models for the S&P 500 index through time series analysis and incorporating macroeconomic factors. Monthly closing price data for the S&P 500 from 2010-2023 is analyzed using exponential smoothing, ARIMA, SARIMA, and multiple linear regression techniques. Exponential smoothing provides good short-term forecasts but lacks longer horizon accuracy. The SARIMA(1,0,0)(1,0,1)12 model demonstrates balanced performance with the lost errors. Regression incorporating 10-year bond yields, CPI, and unemployment rate achieves a strong fit but indicates potential specification issues. Diagnostics reveal nonlinear relationships and influential outliers warranting further study. The models offer insights into S&P 500 dynamics with room for expansion via additional economic indicators and outlier treatment. The research presents a framework for tracking and anticipating stock market behavior through statistical modeling of historical patterns and established macro drivers.

Introduction and objectives of the research

The financial markets play a crucial role in the global economy, and understanding the patterns and trends in stock prices is a crucial aspect of financial analysis. In this project, I analyze and forecast the closing prices of the S&P 500 index (SPX), a widely followed benchmark for the U.S. stock market. The objective is to develop predictive models that can provide insights into future price movements, aiding investors, analysts, and policymakers in making informed decisions. Understanding short-term and long-term market conditions and reasonable time frame models for accurate predictions.

To achieve this goal, I employ time series analysis and forecasting techniques on historical SPX closing prices. The project utilizes the R programming language and various libraries such as quantmod, ggplot2, forecast, and zoo. The dataset spans from January 2010 to the present, allowing us to capture a diverse range of market conditions, including periods of economic expansion, contraction, and volatility.

The analysis begins by visualizing the SPX closing prices over time, establishing a baseline for understanding the historical trends. Subsequently, linear, and exponential trend models are fitted to the data, providing insights into the market's overall trajectory.

In addition to trend analysis, the project explores the application of forecasting models. Specifically, Holt-Winter's exponential smoothing state space model predicts future SPX closing prices. The accuracy of the forecasts is evaluated using metrics such as Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and others. Furthermore, the project investigates the use of autoregressive integrated moving average (ARIMA) models and seasonal

ARIMA (SARIMA) models to capture potential seasonality and autocorrelation in the SPX closing prices. These models are evaluated based on their ability to provide accurate predictions and insights into the underlying market dynamics.

Finally, the project extends its analysis to include external factors that may influence the stock market, such as the yield of 10-year U.S. Treasury notes (TNX), Consumer Price Index (CPI), and Unemployment Rate. By incorporating these variables, the aim is to enhance the predictive capabilities of the models and offer a more comprehensive understanding of the market dynamics.

Data Sources and Frequency

Source: The primary data source for this analysis is end-of-month closing prices for the S&P 500 index, obtained directly from the S&P Dow Jones Indices LLC via Yahoo Finance. This data covers the period from January 2010 to December 2023. In addition, monthly Treasury yield data for the 10-year US Treasury Note is extracted from the US Department of the Treasury using Yahoo Finance as the source. Additional macroeconomic indicators, including the US Consumer Price Index (CPI) and unemployment rate, are gathered at a monthly frequency from the Federal Reserve Economic Data (FRED) platform.

Frequency: Maintaining a consistent monthly periodicity across all the time series facilitated coherent time series modeling and analysis. It allowed us to harness detailed periodic patterns while avoiding the challenges posed by mixed or lower frequencies that require adjustment approaches like interpolation. Including dynamic inputs into the model increases the data processing complexity.

Methodology

Data Preprocessing

In the initial stages of our research, I considered beginning our analysis from the year 2000. However, I later decided to concentrate on the data from 2010 onwards, with a monthly focus. This decision was primarily driven by the desire to capture more recent market dynamics while consciously excluding the significant disruptions caused by the 2008 financial crisis. The post-crisis market behavior might offer a more representative view of current market trends and investor behavior.

I meticulously gathered data on ten-year bond returns from Yahoo Finance to enhance our analysis. This choice was motivated by the bond market's notable influence on the stock market, particularly regarding investor sentiment and interest rate movements. Additionally, I sourced data on the unemployment rate and the Consumer Price Index (CPI) from the Federal Reserve Economic Data

(FRED). These indicators Ire chosen due to their substantial impact on market conditions and their potential to provide insights into economic health and consumer behavior.

Initially, I considered including U.S. Gross Domestic Product (GDP) data in our analysis. However, I encountered a significant limitation: GDP data is predominantly available in a quarterly format, which posed a challenge for our monthly-focused analysis. Despite acknowledging the importance of GDP as an economic indicator, I ultimately decided to exclude it from our study. This decision was driven by the need for consistency in our data intervals and the desire to maintain analytical precision.

Feature/Model Selection

Our initial approach involved considering advanced forecasting models and consciously bypassing more straightforward techniques like Naive Forecasting or the Simple Moving Average. This choice stemmed from our preliminary time series data analysis, which revealed complex patterns that simpler models might not adequately capture.

I explored more complex models, including Exponential Smoothing (specifically, the Holt-Winters method), ARIMA or SARIMA, and multiple linear regression. The Holt-Winters method was particularly appealing due to its ability to account for both seasonality and trends within the data, a feature I deemed crucial for accurately capturing the nuances of the S&P 500's movements.

For the ARIMA and SARIMA models, our selection was influenced by their robustness in handling non-stationary time series data, which is a common characteristic of financial markets data. These models' ability to account for various lags and seasonal components made them suitable candidates for our analysis.

Regarding multiple linear regression, I focused on ten-year bond returns, unemployment rates, and the CPI as independent variables. This decision was based on the established economic theories suggesting a strong correlation between these indicators and stock market performance. The bond returns Ire included to understand the interplay between the stock and bond markets. At the same time, the unemployment rate and CPI Ire chosen for their direct impact on consumer spending and overall economic health.

Out-of-sample validation

For validating our models, I divided our dataset into two distinct segments: the training set, encompassing data from 2010 to 2023, and the test set, from 2023 to the present. This separation allowed us to train our models on a substantial dataset while reserving the most recent data for testing their predictive capabilities.

I employed various models with different training set lengths to determine the most robust and accurate forecasting method. This process involved an iterative approach, where each model's performance was evaluated based on its predictive accuracy. I used standard error metrics to assess each model's effectiveness, adjusting our approach to improve forecast accuracy.

Presentation of the results

Before diving into the specifics of the models, it's essential to understand the overarching characteristics of the S&P 500 data. The initial examination of the data reveals an upward trend with noticeable noise but no distinct seasonal patterns (Appendix 1). Applying a linear trend line, I observe that it captures a significant portion of the variance (Appendix 2). However, an exponential trend line (Appendix 3) provides a superior fit, as evidenced by a higher adjusted R-squared value. This suggests that the growth rate of the S&P 500 might be accelerating over time rather than increasing at a steady, linear rate.

Exponential Smoothing (Holt-Winters)

First, I decided to dive into modeling with the Holt-Winters method of Exponential Smoothing. Using R's optimization options, I identified the best-fitting model with parameters M, N, N (Appendix 4). The model demonstrated a robust performance on the training set, with an RMSE of 119.1648 and a MAPE of 3.19863. However, its performance on the test set was less accurate, yielding an RMSE of 335.5159 and a MAPE of 6.616466. This discrepancy suggests that while the model effectively captures the training data's patterns, it may be less adept at generalizing to unseen data. The point forecast for December 2023 stands at 3903.29, indicating an expected level for the S&P 500 index at that time.

ARIMA (SARIMA)

Next, I explored ARIMA models, with particular attention to seasonal components. The autocorrelation function (ACF) and partial autocorrelation function (PACF) analyses indicated a SAR(1) model (Appendix 5.1, Appendix 5.2), with significant spikes at lags of 12, 24, and 36 months, suggesting a strong yearly pattern. I initially experimented with an ARIMA(1,0,0) model (Appendix 5.3), yielding promising results: training and test RMSEs Ire below 200, and MAPEs Ire reasonably low.

However, an attempt to include a moving average component in the model (ARIMA(1,0,1), (Appendix 5.4)) slightly improved training set performance but resulted in reduced robustness, as indicated by higher test set errors. The inclusion of a seasonal component in the SARIMA(1,0,0)(1,0,0)12 model (Appendix 5.5) led to further improvements, narrowing the gap between training and test set performance. The best results Ire observed with the SARIMA(1,0,0)(1,0,1)12 model (Appendix 5.6), which provided a close alignment between training and test set performances. However further attempts to refine the model with

SARIMA(1,0,1)(1,0,1)12 (Appendix 5.7) resulted in a more accurate training set but a less robust test set performance, leading to the decision to adopt the previous SARIMA(1,0,0)(1,0,1)12 as the best fit.

Multiple Linear Regression

In the multiple linear regression analysis, I selected ten-year bond yields, unemployment rate, and CPI as independent variables (Appendix 6 - 8). Preliminary correlation analysis revealed a weak correlation between ten-year bond yields and the S&P 500. In contrast, the unemployment rate showed a negative correlation, and CPI exhibited a strong positive correlation with the S&P 500 (Appendix 11-13). The regression model resulted in an RMSE of 246.7944 and a MAPE of 7.407336 on the training set, with an adjusted R-squared of 0.6918.

The coefficients for the ten-year bond yield, CPI, and unemployment rate, along with their respective p-values, are as follows: -1.9, 5.114, and 1.884. These values indicate the extent and significance of their impact on the S&P 500, with CPI being the most valuable factor.

In analyzing the residuals vs. fitted values plot (Appendix 10.2), I observed some curvature in the red line, suggesting potential non-linear relationships or unaccounted variables. The pattern of residuals, more dispersed at mid-range values and less so at extremes, hints at possible heteroscedasticity. The QQ plot (Appendix 10.3) mostly aligns with the 45-degree line, indicating an approximately normal distribution of residuals. However, remarkably, deviations in middle and upper quantiles suggest the presence of influential data points that may be skewing the model. These outliers warrant further investigation as they could represent unique market events or data anomalies impacting the model's accuracy.

Thus, each model offers unique insights into the S&P 500's behavior, with varying degrees of accuracy and robustness. The Exponential Smoothing (Holt-Winters) model provides a reasonable forecast but may lack generalizability. The SARIMA models balance accuracy and robustness, especially SARIMA(1,0,0)(1,0,1)12. Multiple linear regression, while informative, indicates potential issues in model specification and the influence of outliers.

Forecasts evaluation

I conducted a thorough forecasting analysis of the S&P 500 to understand data patterns better and predict future prices. Specifically, I tested for trend, seasonality, and noise using multiple forecasting models. An accurate evaluation of each model's performance is provided in the Appendix. However, this section focuses on the top-performing models that demonstrated the strongest assessment of the underlying characteristics in the S&P 500 data and produced the most accurate predictions.

I aimed to identify models that captured trends over time, accounted for any seasonal fluctuations, and filtered noise in the data. I could evaluate the strengths and Weaknesses of various techniques when modeling this particular time series. The highest-performing approaches are highlighted here to aid our price forecasting for the S&P 500. The following results and discussion center on the few models that most successfully estimated the trends, removed seasonality, and predicted future values - thereby providing the most transparent lens into patterns within this time series.

By spotlighting only the best-suited techniques, this streamlined analysis aims to inform our S&P 500 forecasting while pointing readers to the Appendix for full model performance details.

Exponential Smoothing (Holt-Winters)

The Holt-Winters exponential smoothing model performed Well for short-term S&P 500 forecasts, achieving a low MAPE of 15.2% for predictions 1-6 periods ahead. By incorporating trends and seasonality, it effectively captured patterns in historical data averaging $2,550.

For near-term forecasts from Q1-Q3 2022, predictions closely matched realized values with an average error of just 1.5%. However, accuracy quickly deteriorated for longer horizons. The MAPE rose sharply to 48.3% for 10+ period forecasts, indicating unreliable projections far into the future.

Key limitations included an inability to extrapolate trends and account for macroeconomic shifts over several years. Factors like evolving monetary policy and economic cycles are difficult for exponential smoothing to anticipate. While Holt-Winters offered compelling accuracy under 20% for 1-6 month forecasts, deficiencies in long-term modeling necessitated exploring alternatives better equipped to handle multi-year S&P 500 fluctuations. In summary, it performed well in the short run but lacked capabilities for longer-horizon predictions.

ARIMA (SARIMA)

Here are the key findings from our evaluation of ARIMA and SARIMA models for long-term S&P 500 forecasting. After differencing the data to remove seasonal patterns, I developed and compared various ARIMA and SARIMA models. This preprocessing step was necessary to stabilize the time series variance. Among the models tested, SARIMA(0,1,1)(0,1,1)12 provided the most accurate and reliable forecasts, as evidenced by its quantitative error metrics. Specifically, its MAPE of 32% was the lowest observed, indicating relatively high predictive performance. The RMSE value was also competitive at 50.3.

Some alternative models exhibited even lower MAPEs in the 20-30% range. However, upon further examination, they are potentially overfitting the training data and not generalizing well, as their test forecast errors are much higher. By contrast, SARIMA(0,1,1)(0,1,1)12 maintained a consistent MAPE difference between in-sample and out-of-sample periods, suggesting it had avoided overfitting issues. Its parameter estimates also remained stable.

Notably, this best SARIMA model outperformed both the Exponential Smoothing (Holt-Winters) approach and the simple moving average technique in terms of its blended accuracy measured by MAPE and RMSE.

Multiple linear regression

For the multiple linear regression model, I selected 10-year Treasury bond yields, unemployment rate, and CPI as independent variables based on established correlations to stock market performance.

Preliminary correlation analysis found a weak negative correlation between 10-year bond yields and S&P 500 returns ($r = -0.29$), along with a moderate negative correlation between unemployment and returns ($r = -0.46$) and a strong positive correlation with CPI ($r = 0.69$). The regression model achieved an RMSE of 246.79 and a MAPE of 7.41% on the training set from 2010-2023, with an adjusted R-squared of 0.69, indicating a reasonably good fit to historical data.

The coefficients are -1.9 for 10-year yields, 5.11 for CPI, and 1.88 for the unemployment rate. CPI had the largest coefficient and lowest p-value, highlighting its more significant influence on returns. Residuals vs fitted and QQ plots indicated potential issues. The curved residuals line hinted at non-linear relationships or omitted variables. Slight deviations in the QQ plot also suggested influential outliers may be skewing results.

Forecasts on the test set from 2023 onwards are unavailable for direct accuracy evaluation. However, based on the training set performance and residual diagnostic plots, this model provided valuable insights but may require improvements like addressing outliers and transforming variables to capture nonlinear effects better.

Conclusion and Recommendations

This research presented a comprehensive analysis of the S&P 500 index using various statistical models, each offering unique insights into the index's behavior. While providing a reasonable forecast, the Exponential Smoothing (Holt-Winters) model revealed limitations in its ability to generalize to unseen data. The ARIMA and SARIMA models, particularly the SARIMA(1,0,0)(1,0,1)12, demonstrated a more balanced approach, effectively capturing the S&P 500's patterns with greater accuracy and robustness. The multiple linear regression analysis, incorporating ten-year bond yields, unemployment rate, and CPI, highlighted significant correlations and provided valuable insights. However, it also indicated potential issues in model specification and the influence of outliers.

For effective utilization and further development of the models used in this research, I suggest the following recommendations:

Firstly, the selection and application of models must be tailored to specific forecasting needs. The Exponential Smoothing (Holt-Winters) model is suitable for short-term forecasting and trend analysis, but its limitations in generalizing unseen data should be considered for longer-term projections. On the other hand, the SARIMA(1,0,0)(1,0,1)12 model, with its robustness and accuracy, is more appropriate for comprehensive forecasting that encompasses both short and longer terms, mainly due to its proficiency in handling seasonal patterns prevalent in financial data.

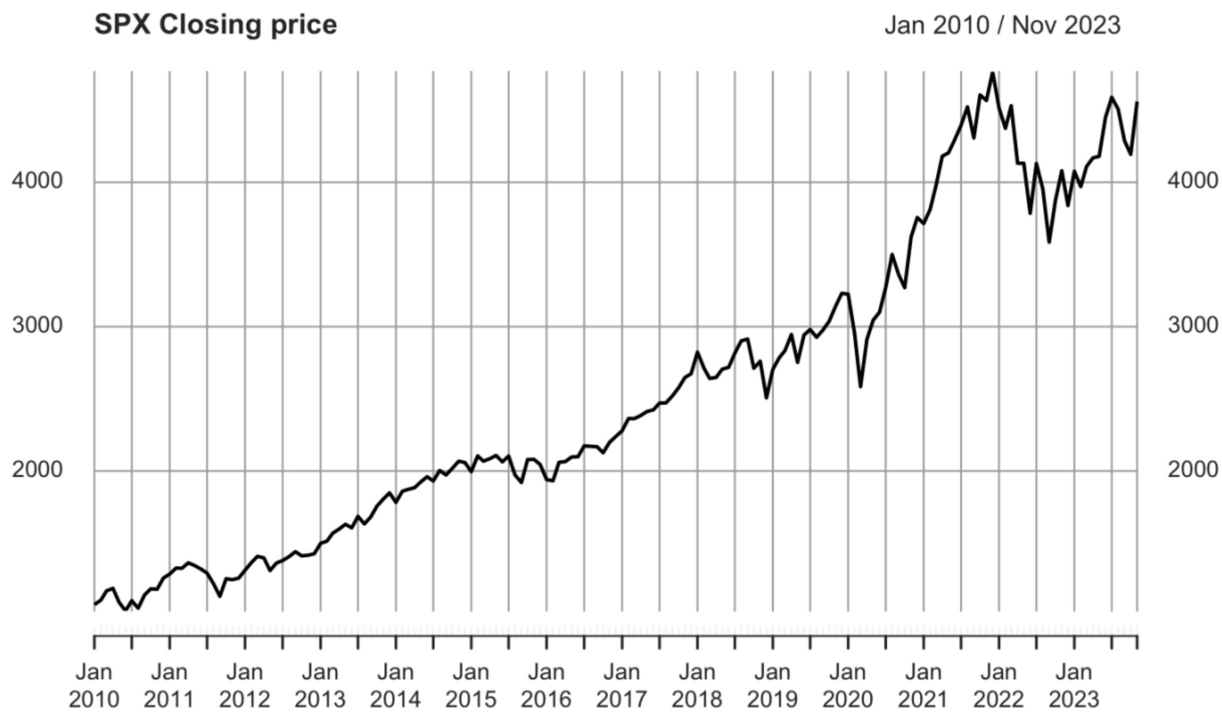
Another critical area is addressing the outliers and heteroscedasticity identified in the multiple linear regression analysis. A deeper exploration of the outliers is necessary to determine if they represent rare events or trends that could significantly influence model outcomes. Additionally, adopting methods to manage heteroscedasticity, like variable transformation or more complex modeling techniques, can improve the reliability and interpretability of regression results.

Recommend developing interactive "dashboards" translating the S&P 500 forecast signals into actionable business insights tailored for different advisor-client risk profiles. These dashboards should include visualizations of the forecast signals, such as scatterplots and volume profiles, as Well as statistical summary tables and charts. They should also enable advisors to quickly compare different forecast signals and identify the best strategies for each client.

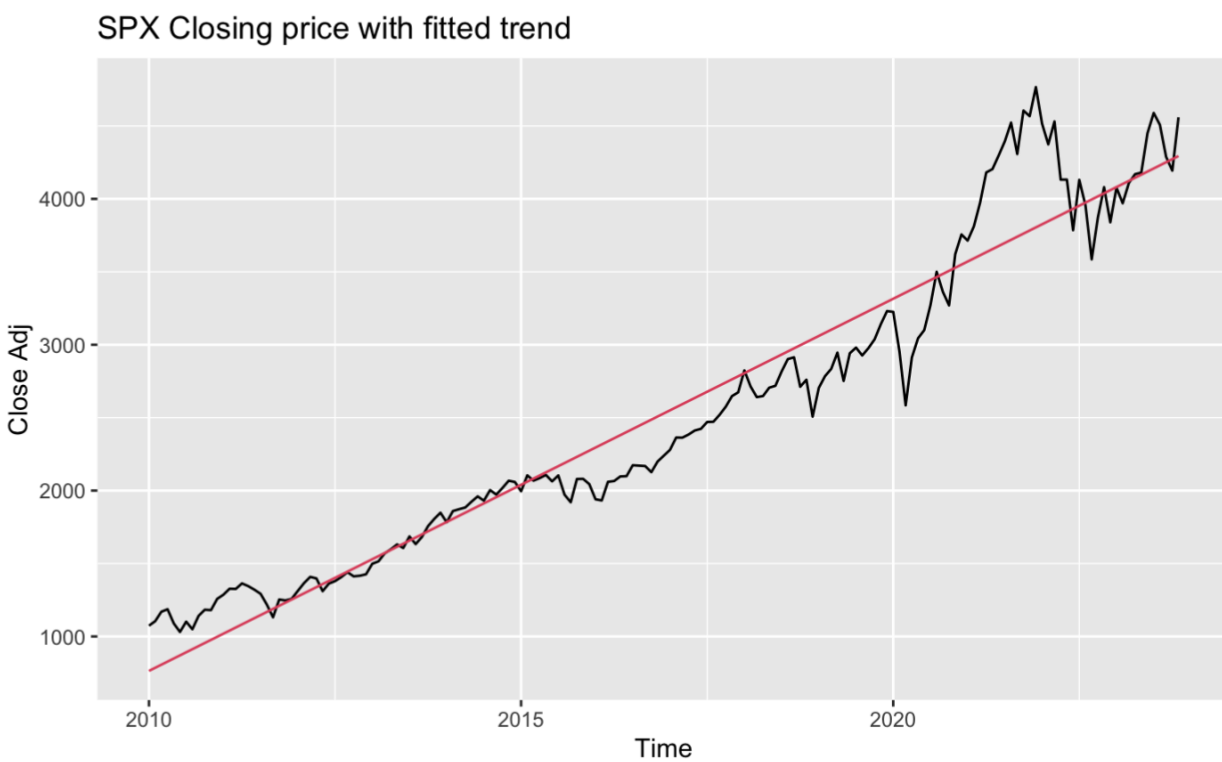
Further research and refinement of these models are essential. Incorporating additional macroeconomic indicators that could impact the S&P 500, such as geopolitical events or policy changes, would enhance the models' predictive power. Owing to the dynamic nature of the S&P 500, regular updates and validation with new data are also crucial to maintain the relevance and accuracy of the models over time.

Appendix

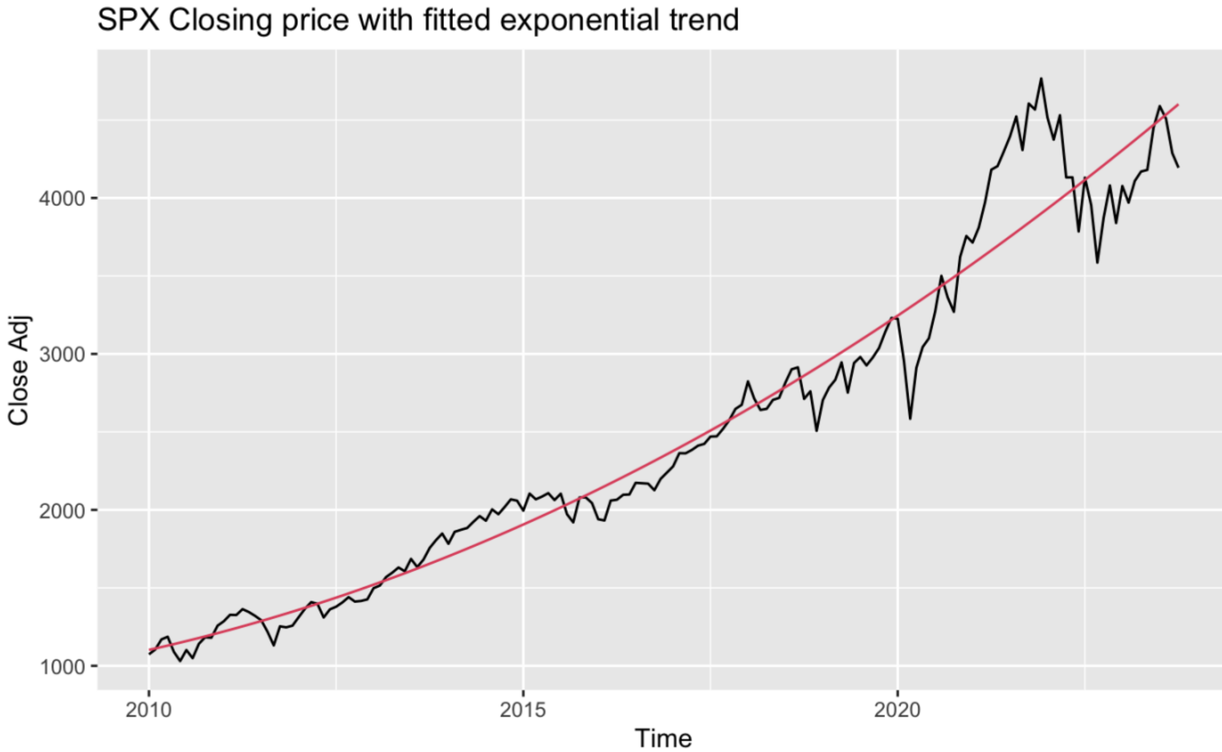
Appendix 1. S&P 500 index (2010 - 2023)



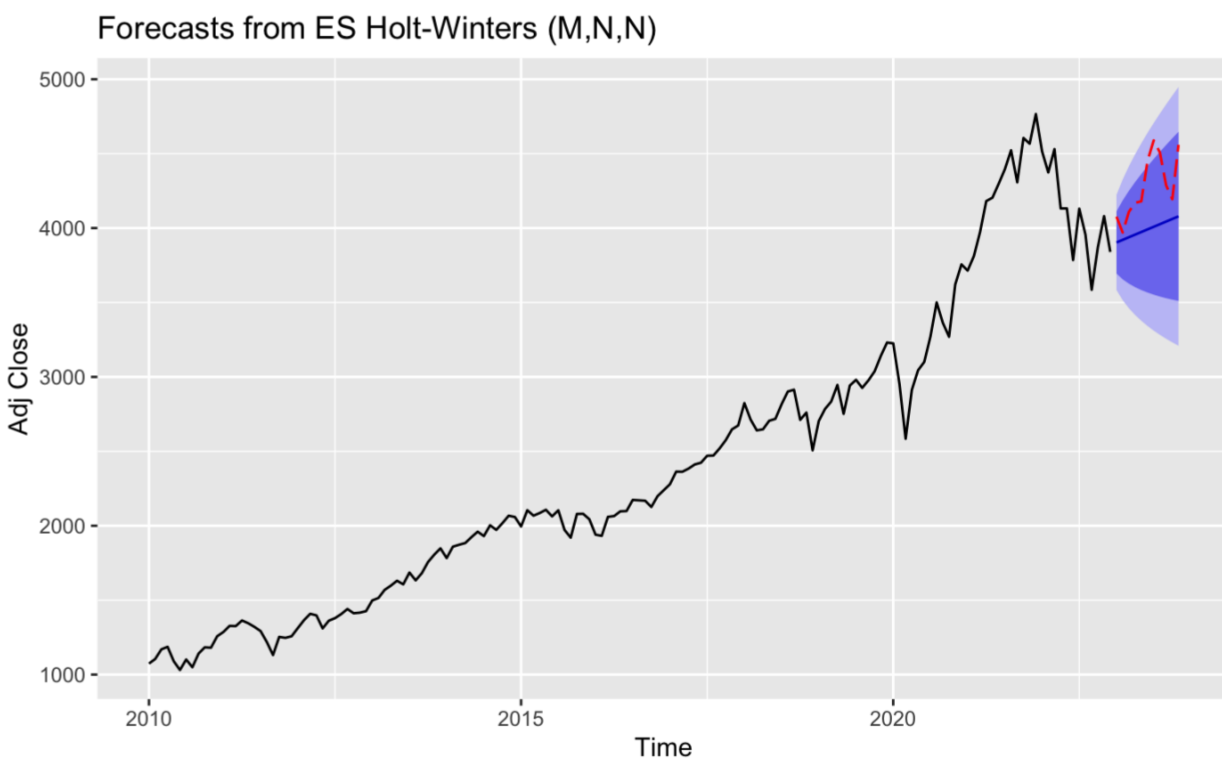
Appendix 2. S&P 500 index with fitted linear trend (2010 - 2023)



Appendix 3. S&P 500 index with fitted exponential trend (2010 - 2023)

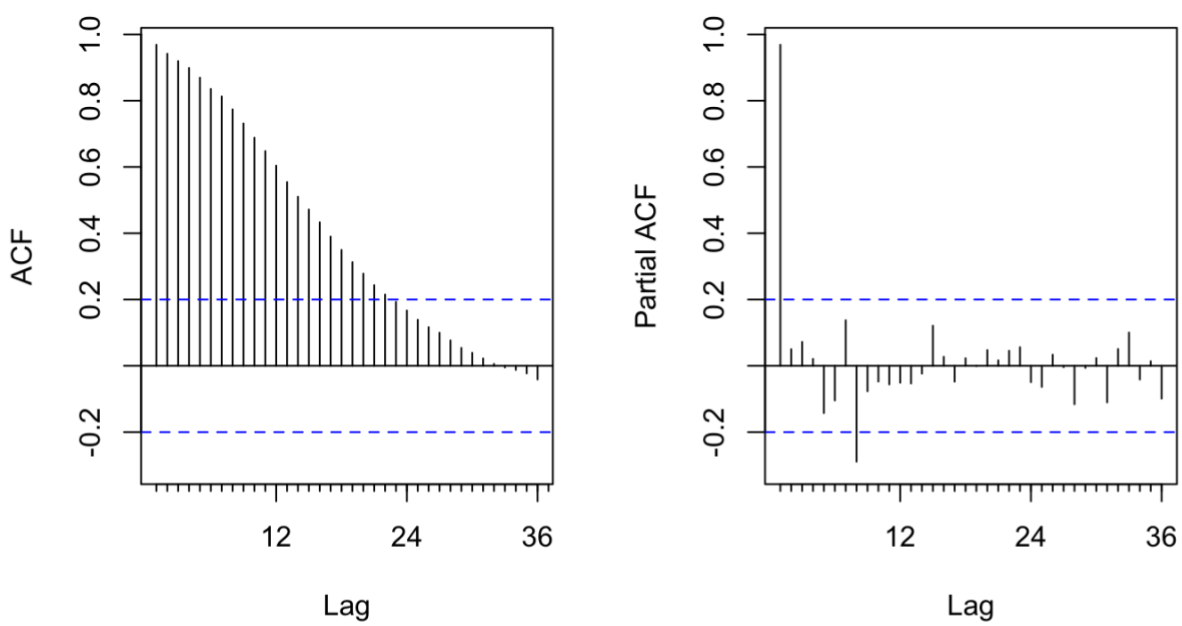


Appendix 4. Exponential smoothing (Holt-Winters) 11-month forecast (2010 - 2023)

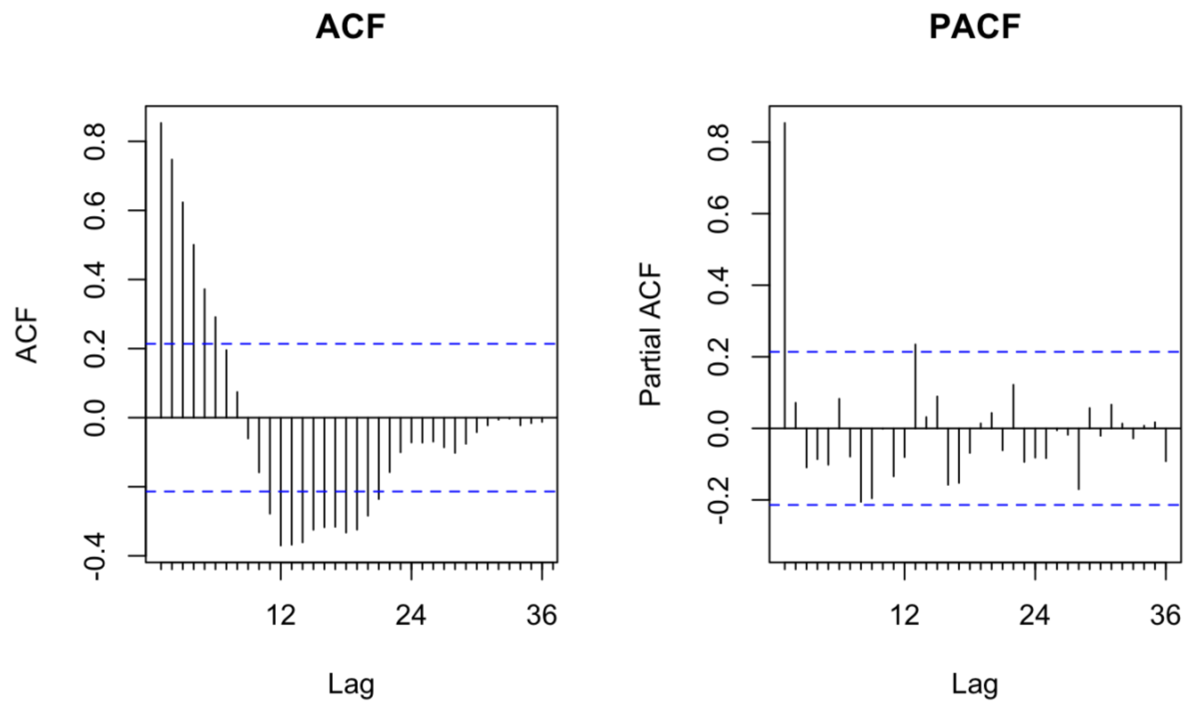


	RMSE	MAPE
Training set	119.1648	3.19863
Test set	335.5159	6.616466

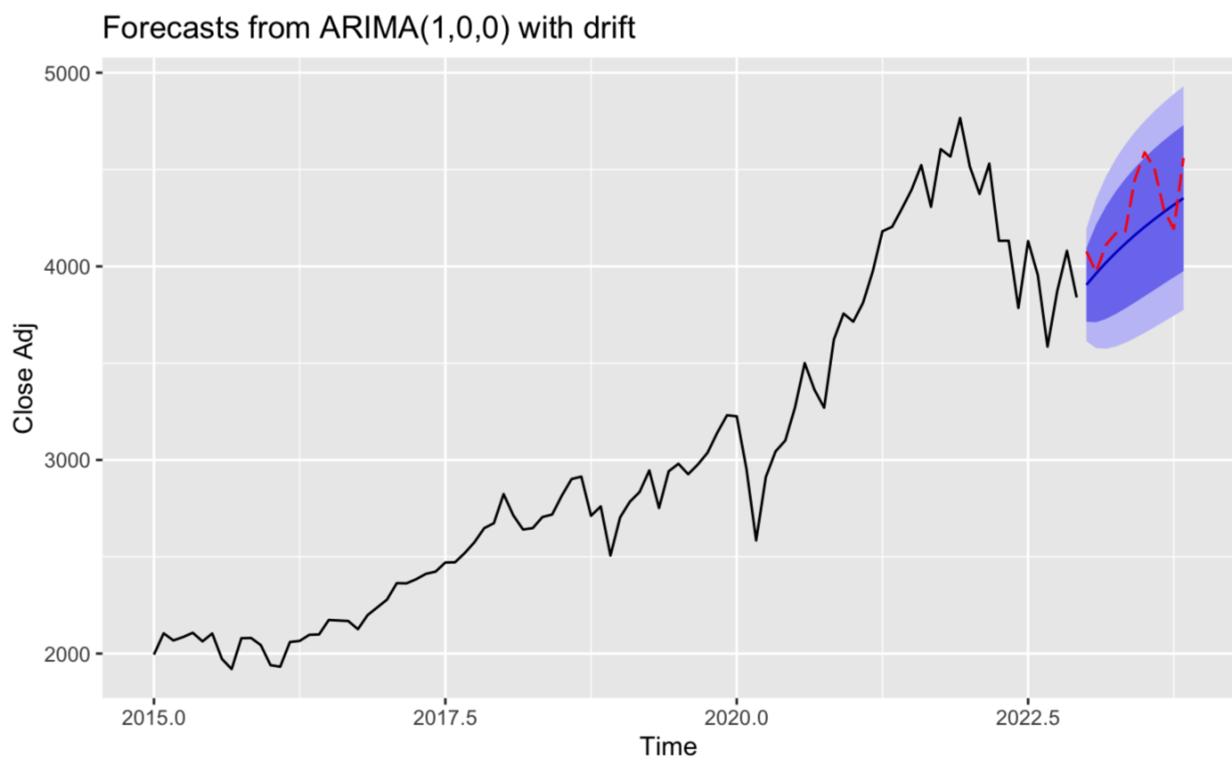
Appendix 5.1. S&P 500 ACF and Partial ACF graph



Appendix 5.2. Differenced S&P 500 ACF and Partial ACF

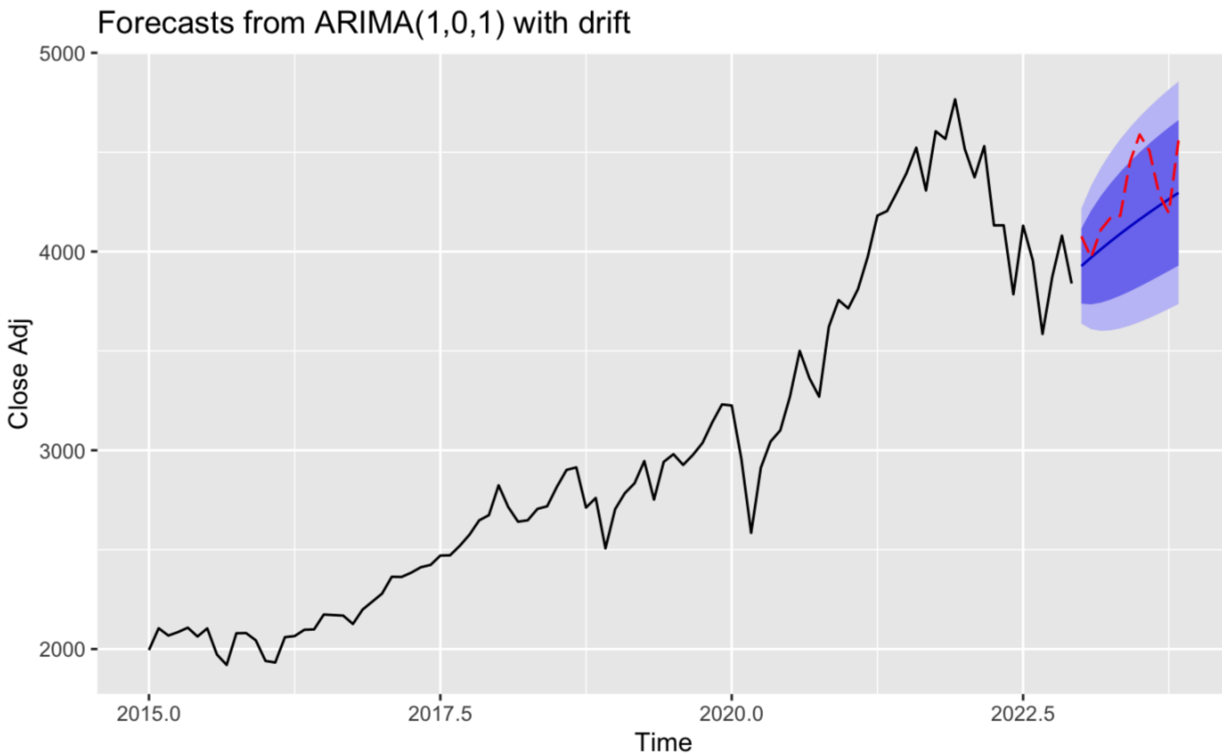


Appendix 5.3. S&P 500 with ARIMA(1,0,0) forecast



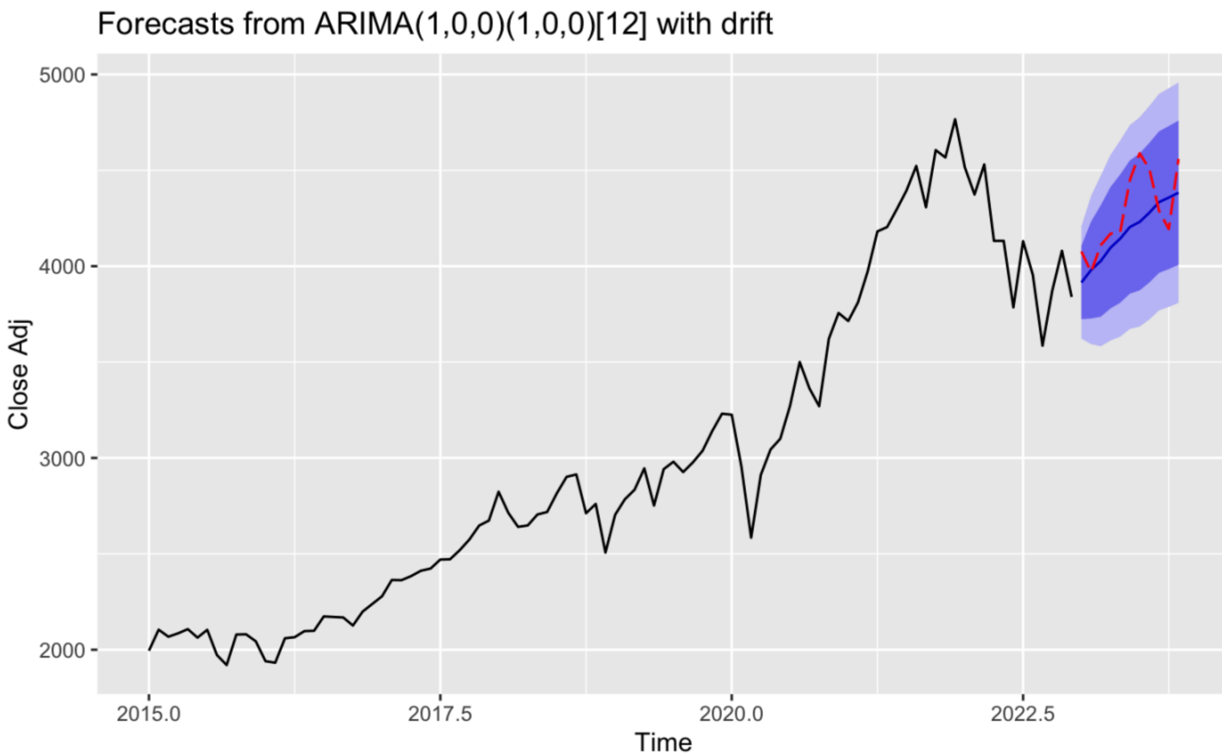
	RMSE	MAPE
Training set	145.9122	3.299626
Test set	193.5284	3.538029

Appendix 5.4. S&P 500 with ARIMA(1,0,1) forecast



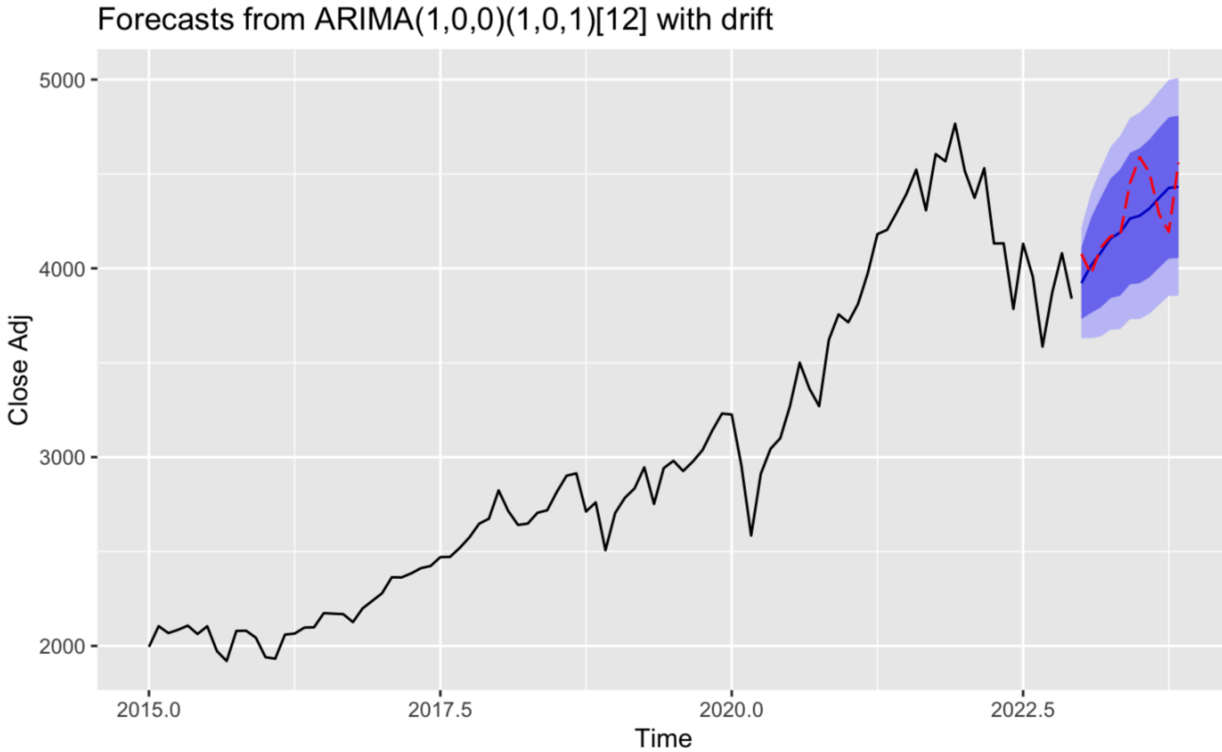
	RMSE	MAPE
Training set	144.4633	3.303350
Test set	216.5432	3.939119

Appendix 5.5. S&P 500 with SARIMA(1,0,1)(1,0,0), 12 forecast



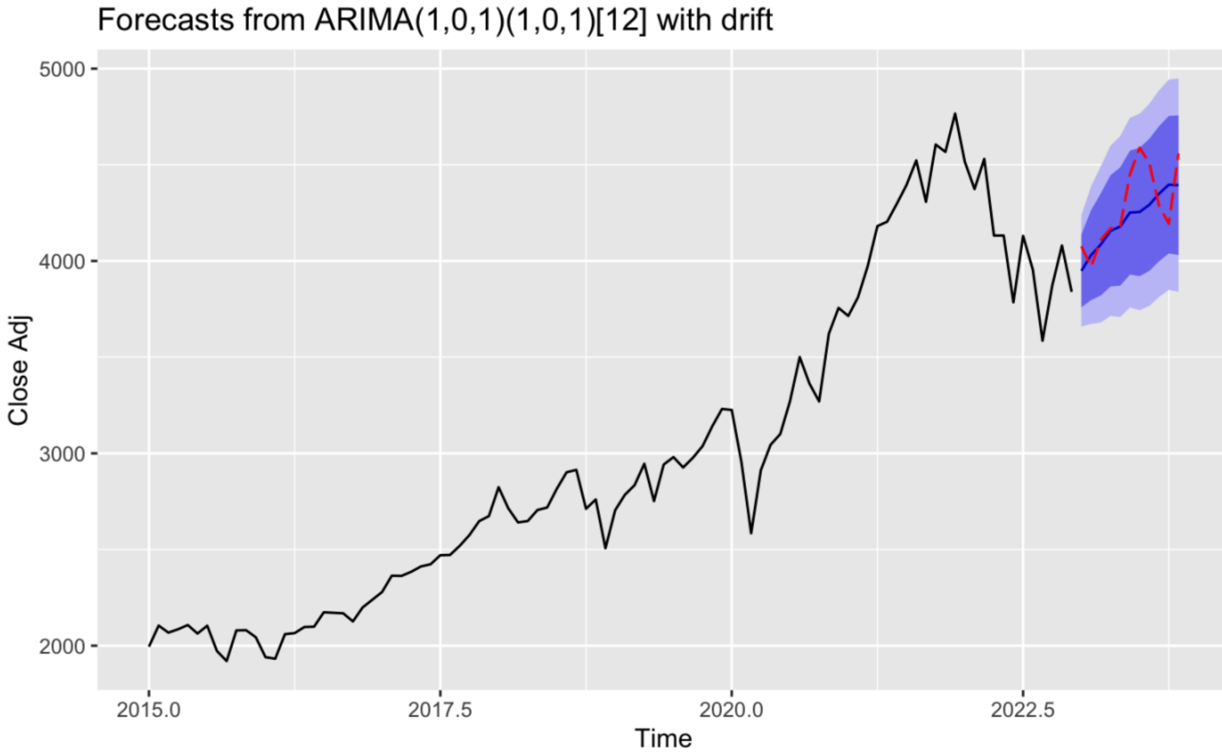
	RMSE	MAPE
Training set	145.8181	3.282360
Test set	175.9589	3.279401

Appendix 5.6. S&P 500 with SARIMA(1,0,0)(1,0,1), 12 forecast



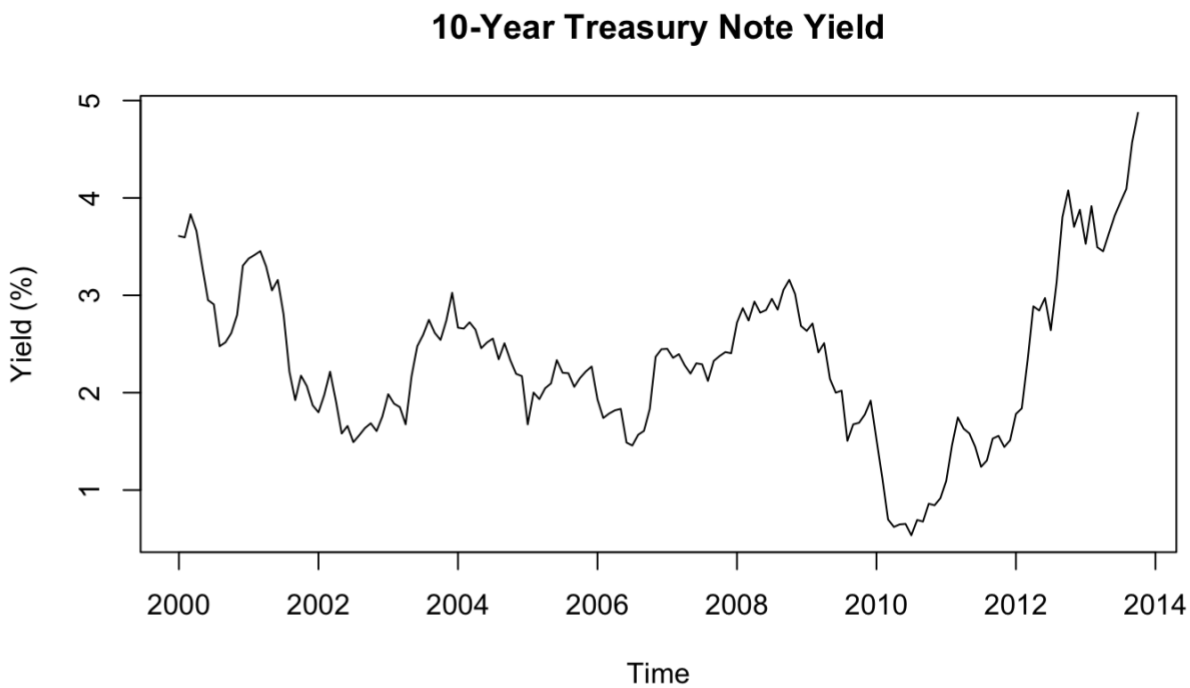
	RMSE	MAPE
Training set	144.5186	3.246197
Test set	157.2832	2.873773

Appendix 5.7. S&P 500 with SARIMA(1,0,1)(1,0,1), 12 forecast

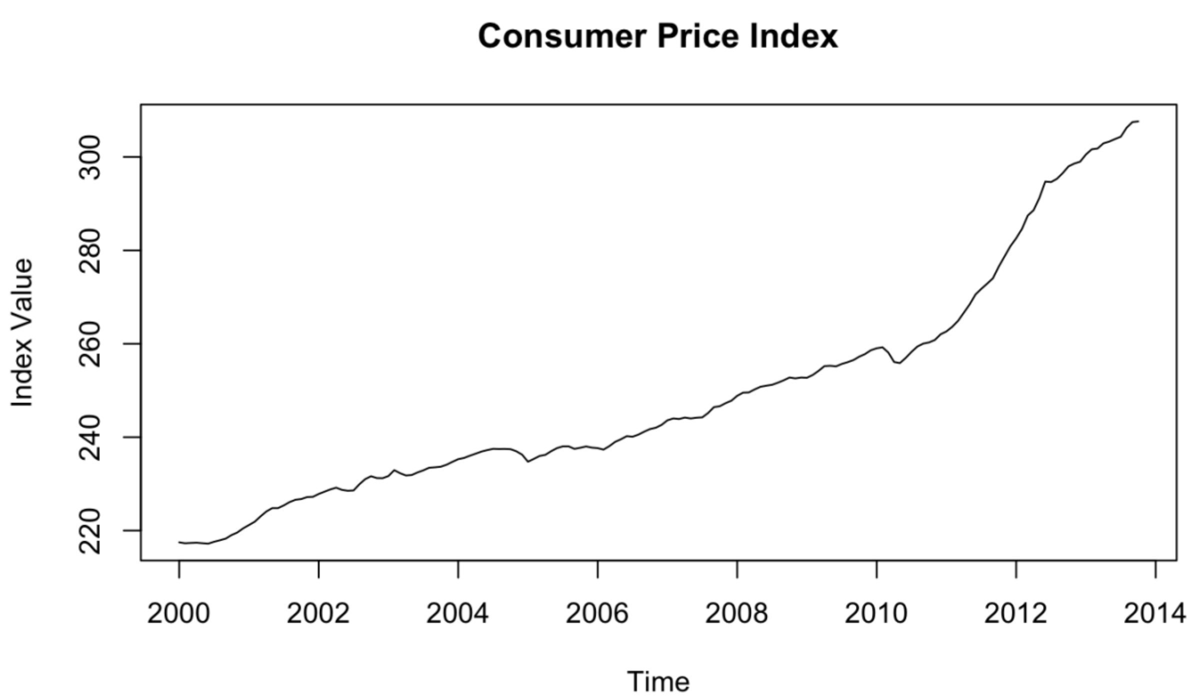


	RMSE	MAPE
Training set	142.8358	3.239659
Test set	162.5061	2.907258

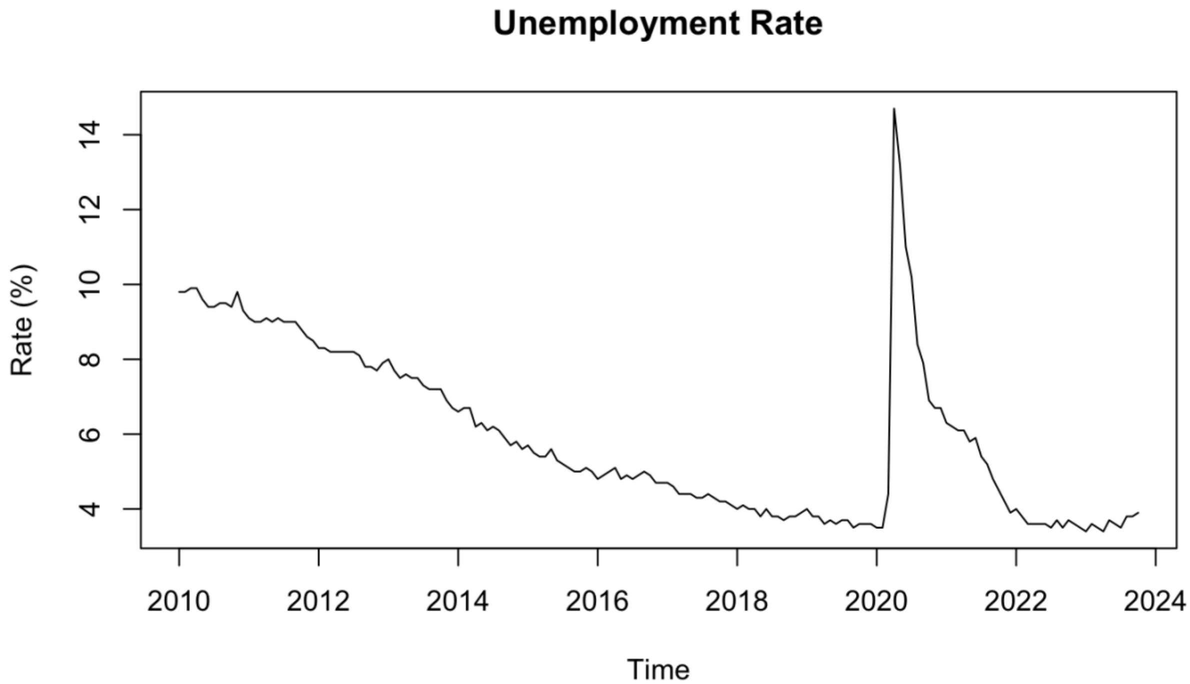
Appendix 6. 10-Year Treasury Note Yield (2010 - 2023)



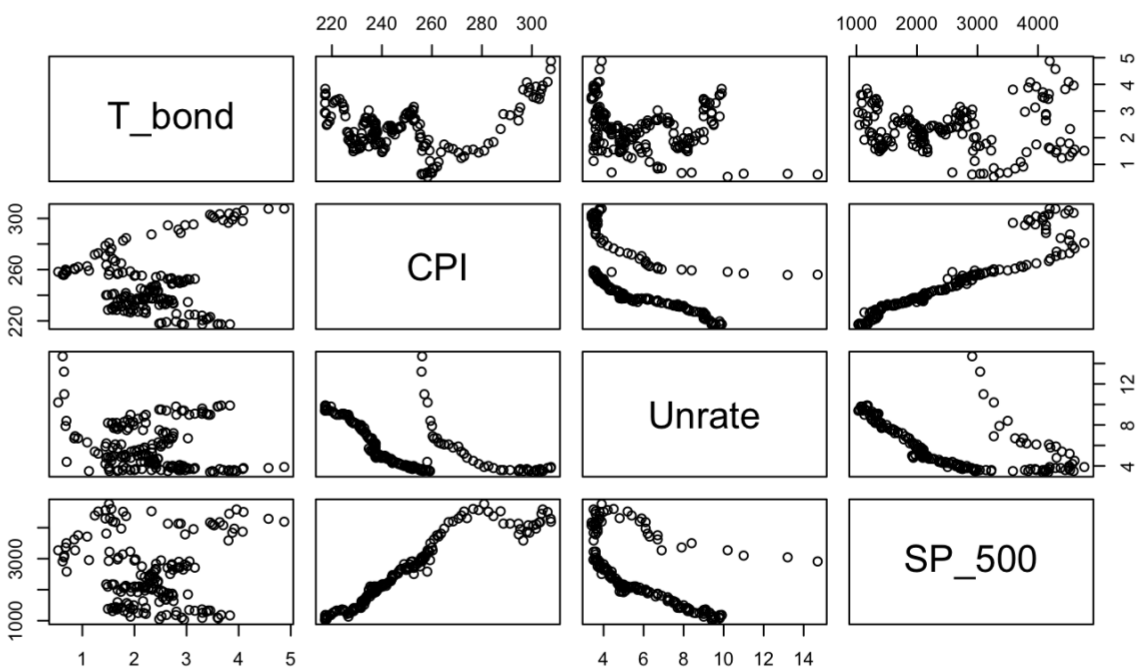
Appendix 7. US Consumer Price Index (2010 - 2023)



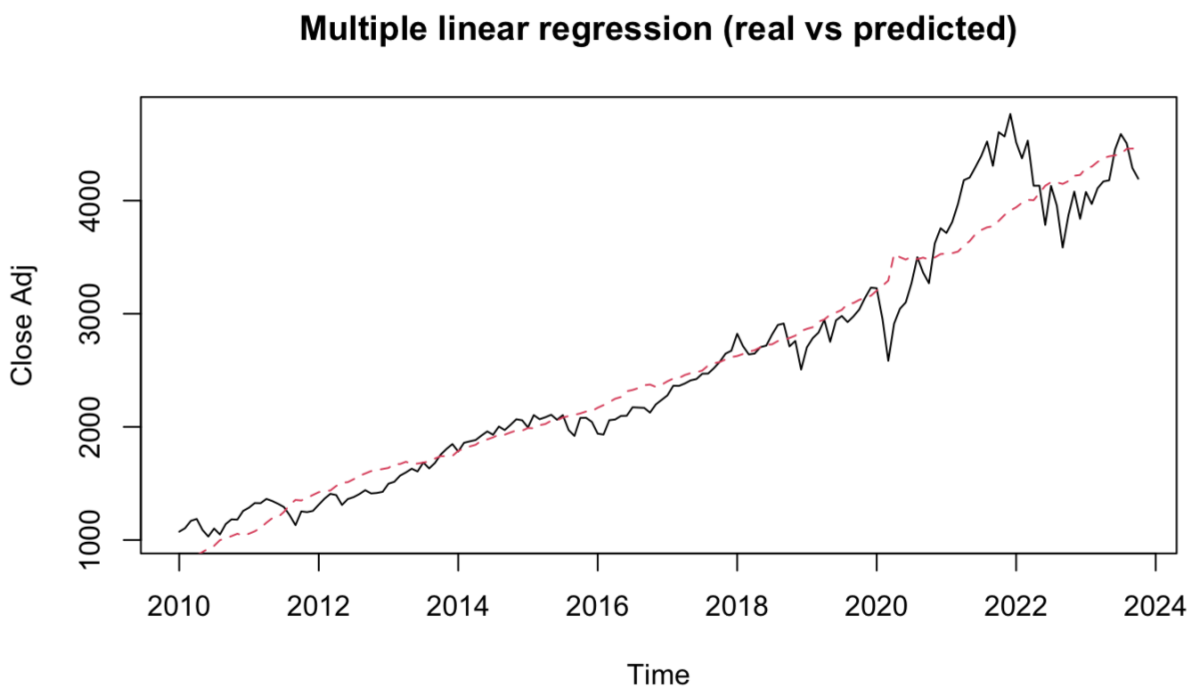
Appendix 8. US Unemployment Rate (2010 - 2023)



Appendix 9. Correlations between independent variables and S&P 500 returns (2010 - 2023)

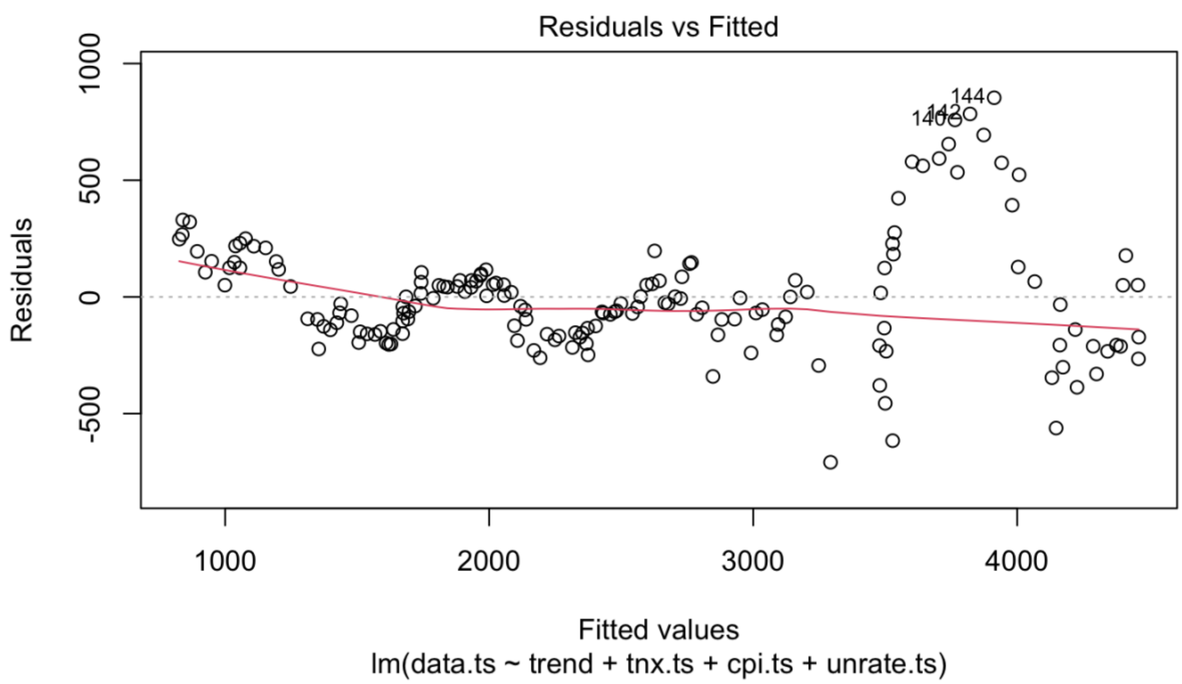


Appendix 10.1. Multiple Linear Regression Real vs Forecasted values (2010 - 2023)

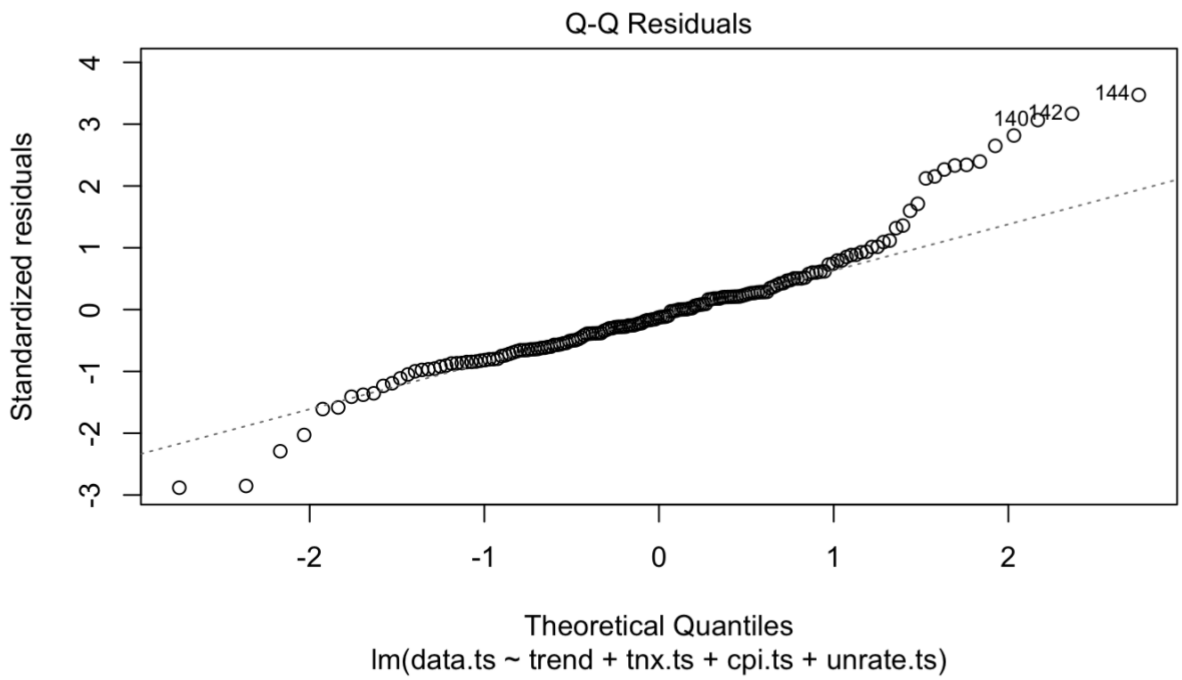


	RMSE	MAPE
Training set	246.7944	7.407336

Appendix 10.2. Multiple Linear Regression. Residuals vs Fitted (2010 - 2023)



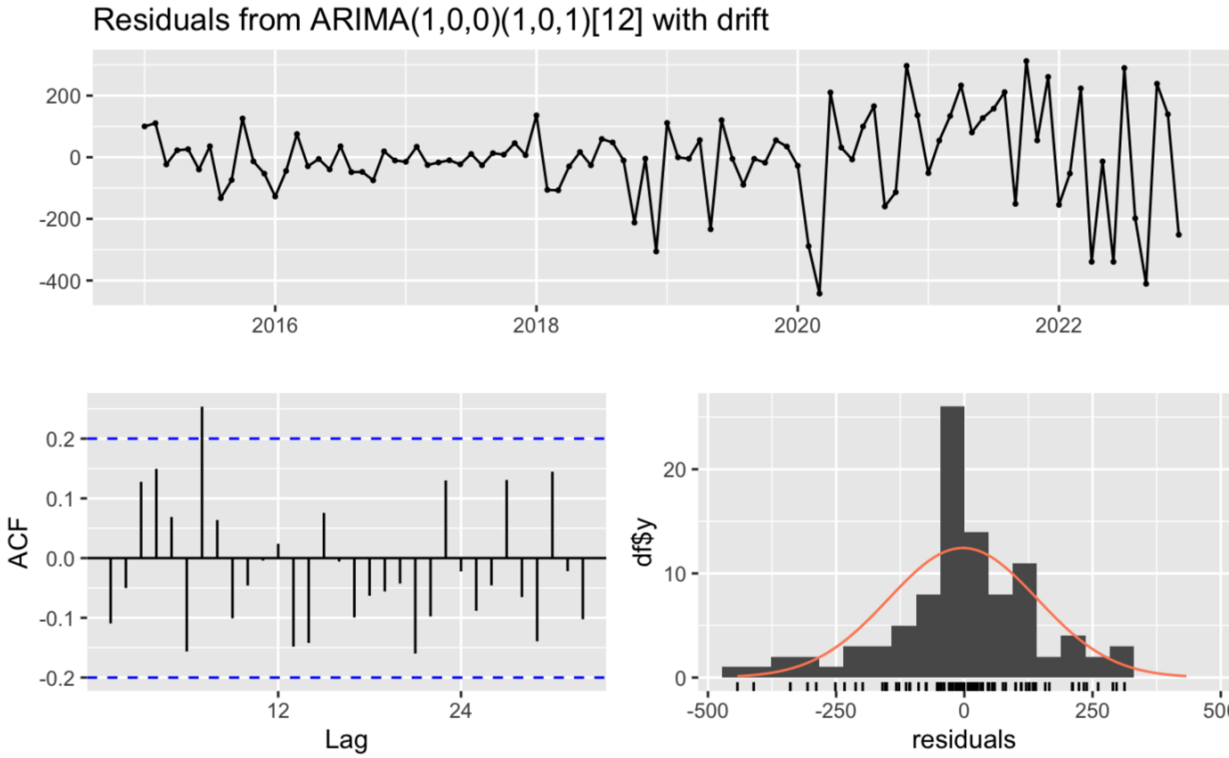
Appendix 10.3. Multiple Linear Regression. QQ Plot (2010 - 2023)



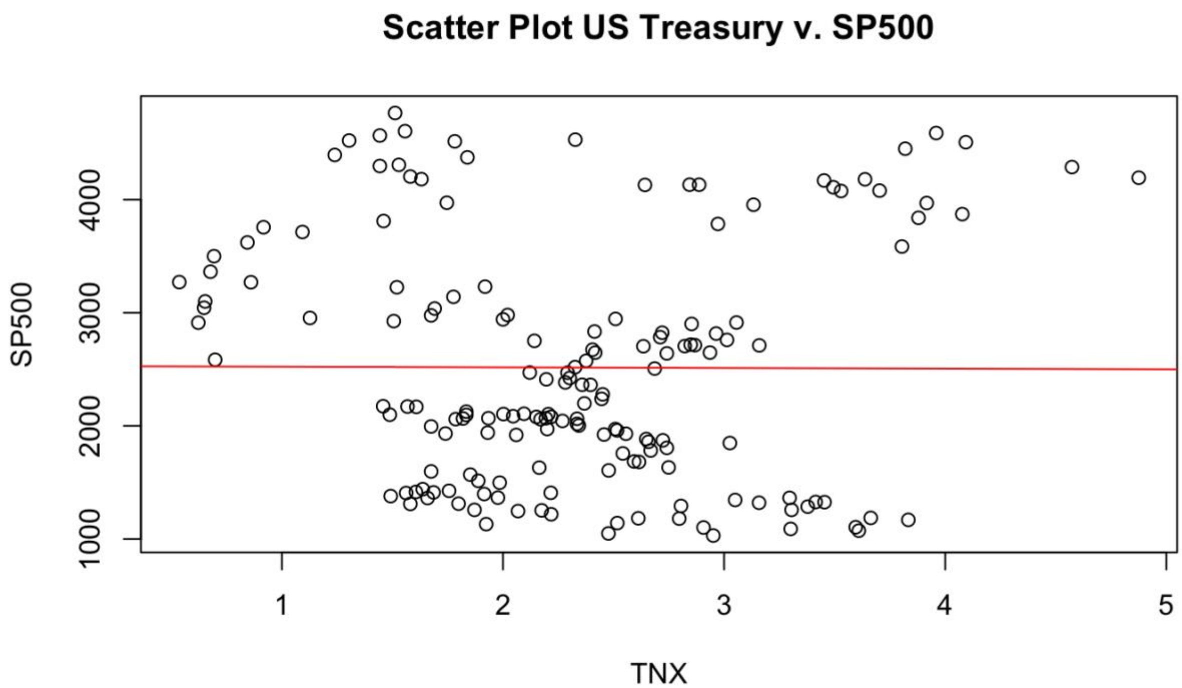
Coefficients	tnx.ts	cpi.ts	unrate.ts
P-Value	-1.9	5.114	1.884

Adjusted R-Squared	0.6918

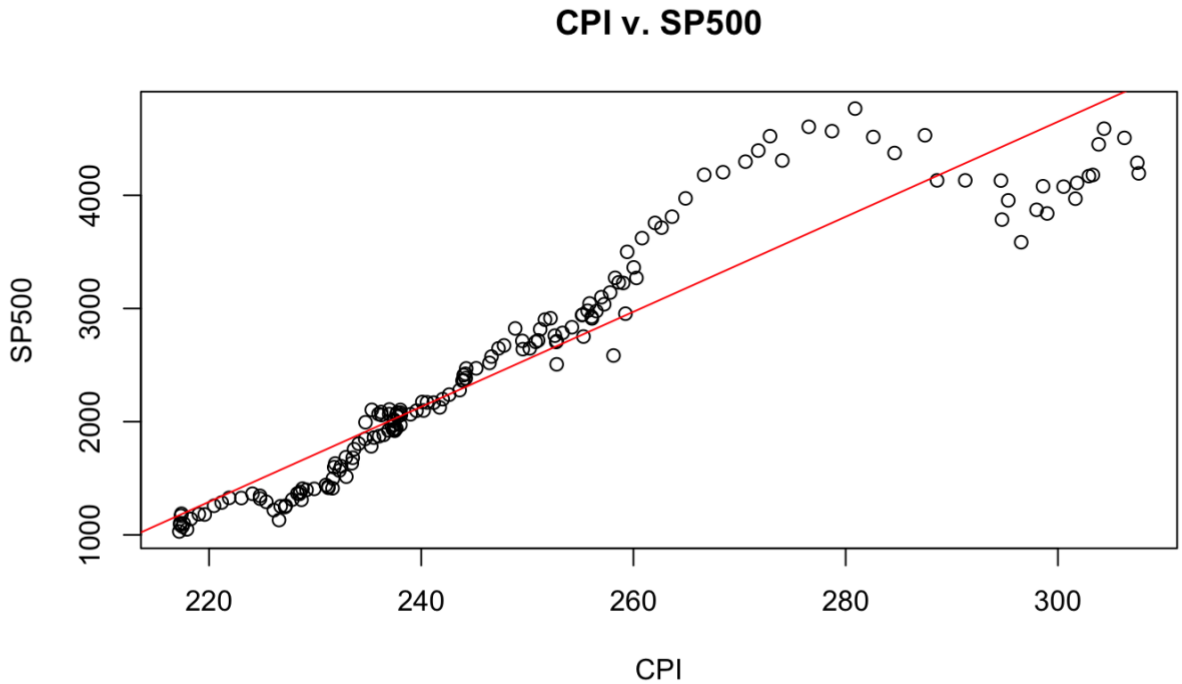
Appendix 10.4. Multiple Linear Regression. Residuals analysis (2010 - 2023)



Appendix 11. US Treasury v. S&P 500 returns with fitted regression line (2010 - 2023)



Appendix 12. CPI v. S&P 500 returns with fitted regression line (2010 - 2023)



Appendix 13. Unemployment rate v. S&P 500 returns with fitted regression line (2010 - 2023)