Time Series Analysis: Forecasting SPY Daily Returns using ARIMA
by Jonathan Jaimangal

1. Introduction

1.1 Question of Interest

The SPY is an exchange-traded fund (ETF) which aims to track the performance of the S&P 500 index. The S&P 500 index tracks the top 500 publicly traded companies in the US (weighted by market cap). It is one of the most actively traded ETFs and has been on an uptrend until 2020. The aim of this report is to compare a trading strategy from an ARMA model, with non-statistical baseline models.

1.2 The Data

The Adjusted Close data used in this project was obtained from the Yahoo Finance API in Python. Daily returns (which are the focus of this report) were calculated based on the Adjusted Close price. A test set of 10 years of daily returns is used to forecast predictions for the following year.

In this report, the Augmented Dickey-Fuller test is used to determine stationarity. When applied to the Adjusted Close data, the p-value is 0.99. Since the p-value is not less than 0.05, we fail to reject the null hypothesis and this time series is non-stationary. Similarly, for the Daily Change data, the p-value is approximately 0 indicating this time series is stationary.

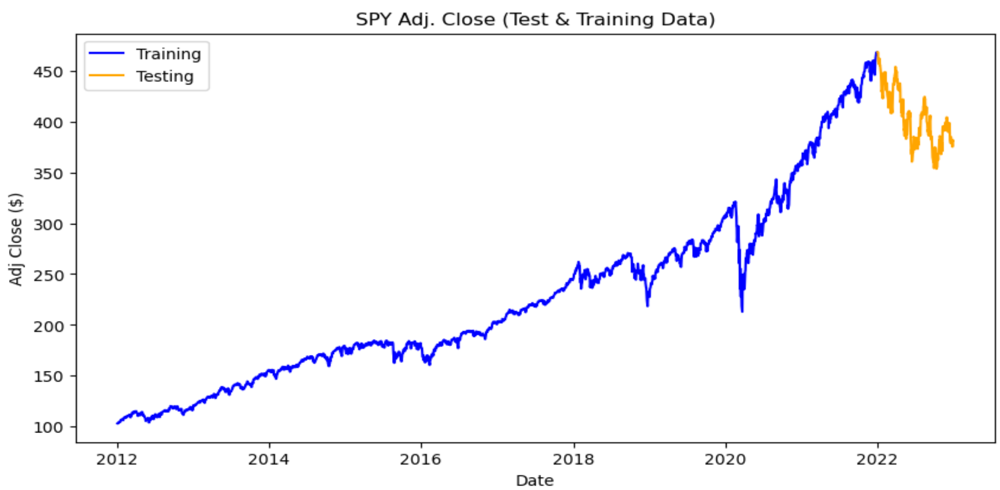

Figure 1. Daily Adjusted Close of the SPY from 1/1/2012 to 1/2/2023

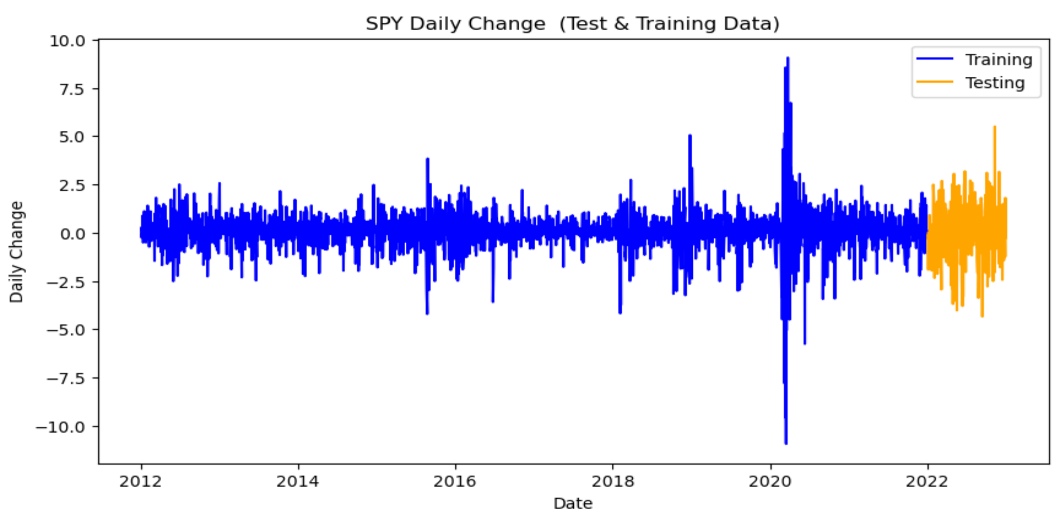

Figure 2. Daily Change of the SPY from 1/1/2012 to 1/2/2023

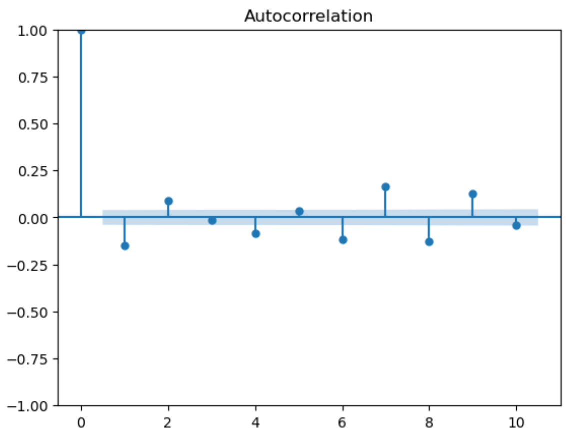

Figure 3. ACF Plot of Daily Change

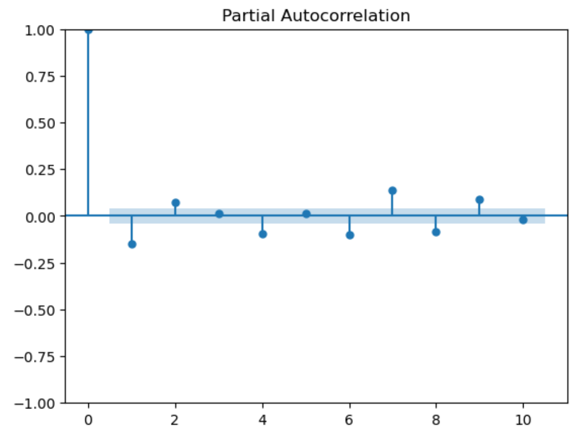

Figure 4. PACF Plot of Daily Change

1.3 The Outliers

During this period, there were several outliers worth noting. On 2/5/2018, the largest single day loss was recorded and on 12/26/2018, the largest single day gain. On 2/19/2020, the S&P 500 reached its highest point during the bull market. On 3/16/2020, the S&P suffered one of the largest single day decreases in history due to the Coronavirus pandemic. On 12/29/2021, the S&P 500 closed at a record high. These were reflected in the Adjusted Close data of the SPY.

2. Methods

2.1 Model Specification

As shown in figures 3 and 4 there were several significant lags. Based on the Akaike Information Criteria (AIC), the best model to forecast daily SPY returns is ARIMA (6,0,2).

p	d	q	AIC
6	0	2	7028
7	0	3	7028
6	0	3	7030
7	0	2	7030
6	0	4	7031
5	0	4	7037
7	0	1	7038
6	0	2	7040
4	0	4	7044
4	0	3	7045

Figure 5. AIC calculations for various ARIMA models

2.2 Model Fitting

Our (6,0,2) model can be represented by the following equation:

(1) Yt = -1.4441(Yt-1) - 0.678(Yt-2) + 0.0219(Yt-3) - 0.0402(Yt-4) - 0.1181(Yt-5) - 0.133(Yt-6) + 1.3444(εt−1) + 0.5932(εt−2) + 0.0651

```
                               SARIMAX Results
==============================================================================
Dep. Variable:           Daily Change   No. Observations:                 2517
Model:                 ARIMA(6, 0, 2)   Log Likelihood               -3504.190
Date:                Sat, 20 May 2023   AIC                           7028.380
Time:                        19:33:34   BIC                           7086.689
Sample:                             0   HQIC                          7049.542
                               - 2517
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0651      0.019      3.354      0.001       0.027       0.103
ar.L1         -1.4441      0.053    -27.347      0.000      -1.548      -1.341
ar.L2         -0.6780      0.057    -11.818      0.000      -0.790      -0.566
ar.L3          0.0219      0.020      1.079      0.281      -0.018       0.062
ar.L4         -0.0402      0.019     -2.147      0.032      -0.077      -0.004
ar.L5         -0.1181      0.019     -6.242      0.000      -0.155      -0.081
ar.L6         -0.1330      0.014     -9.456      0.000      -0.161      -0.105
ma.L1          1.3444      0.052     25.896      0.000       1.243       1.446
ma.L2          0.5932      0.047     12.739      0.000       0.502       0.685
sigma2         0.9488      0.012     77.205      0.000       0.925       0.973
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):             13377.40
Prob(Q):                              0.98   Prob(JB):                         0.00
Heteroskedasticity (H):               2.79   Skew:                            -0.95
Prob(H) (two-sided):                  0.00   Kurtosis:                        14.13
===================================================================================
```

Figure 6. Summary of the chosen ARIMA model

2.3 Model Diagnostics

We expect the residuals to be independently and identically distributed normal variables with zero mean and constant variance. In this report, residuals will be analyzed using a Quantile-Quantile plot, Density Histogram of Residuals and the Shapiro-Wilk normality test. Looking at figure 7, the residuals seem to be random with zero mean. The density plot in figure 8, appears to be normally distributed with a mean of zero. The Quantile-Quantile plot in Figure 9 is approximately linear implying that the error terms are normally distributed. The p-value of the Shapiro-Wilk test is very close to 1, which is above 0.05 indicating the residuals are normally distributed.

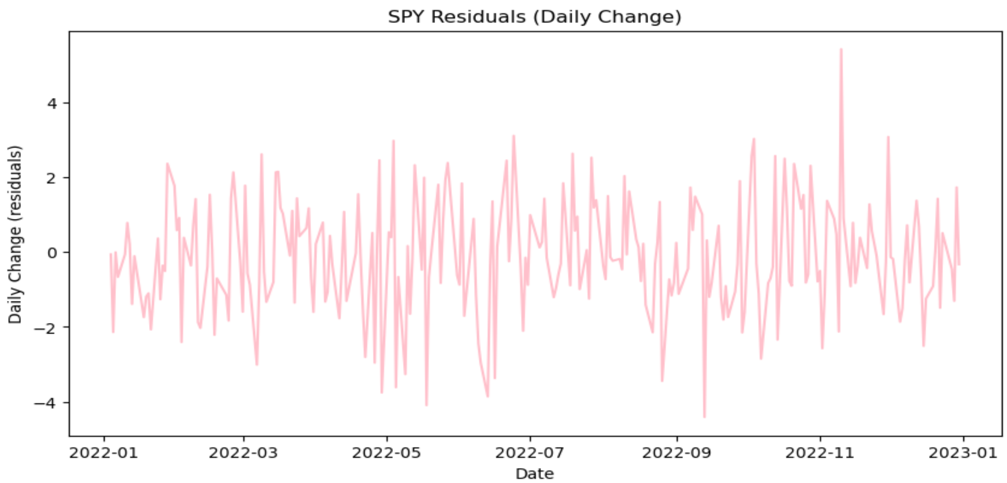

Figure 7. Plot of the residuals from ARIMA (6,0,2)

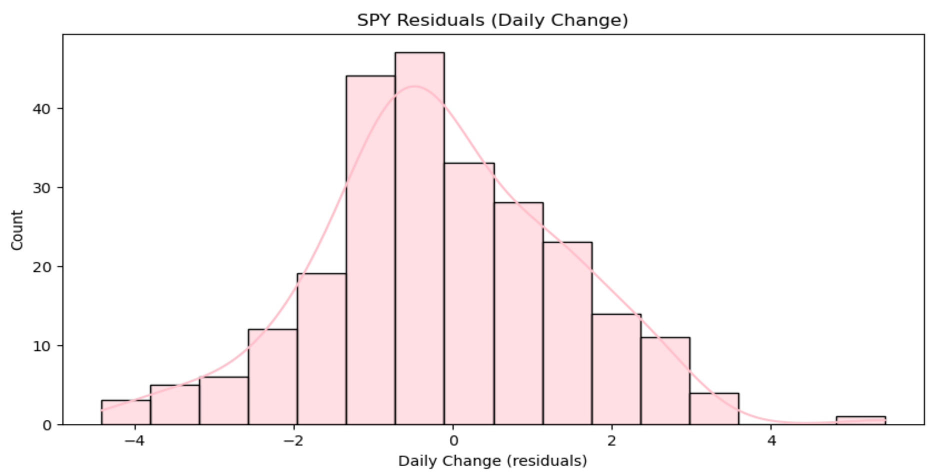

Figure 8. Density of residuals from ARIMA (6,0,2)

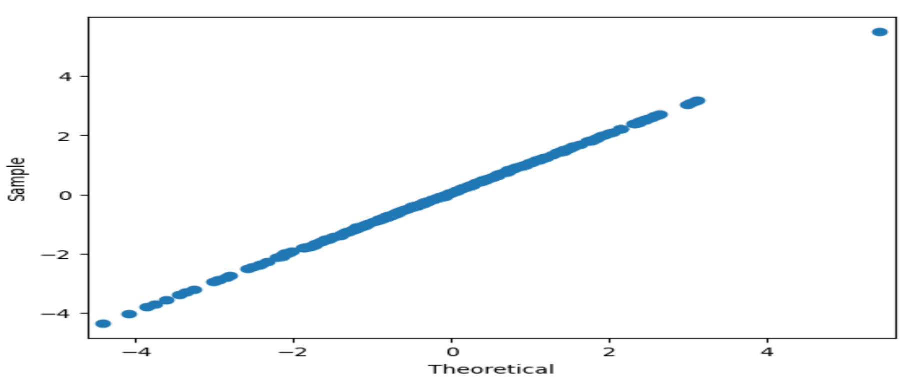

Figure 9. QQ Plot of the ARIMA (6,0,2)

3. Results

3.1 Forecast Results

The forecast suggests some price fluctuations from Jan-2022 to Mar-2022 followed by a linear increase thereafter. The actual daily changes fluctuate much more and the actual adjusted close declines during this period. The forecasts are positive when the actual changes are positive or negative when the actual changes are negative about 45% of the time.

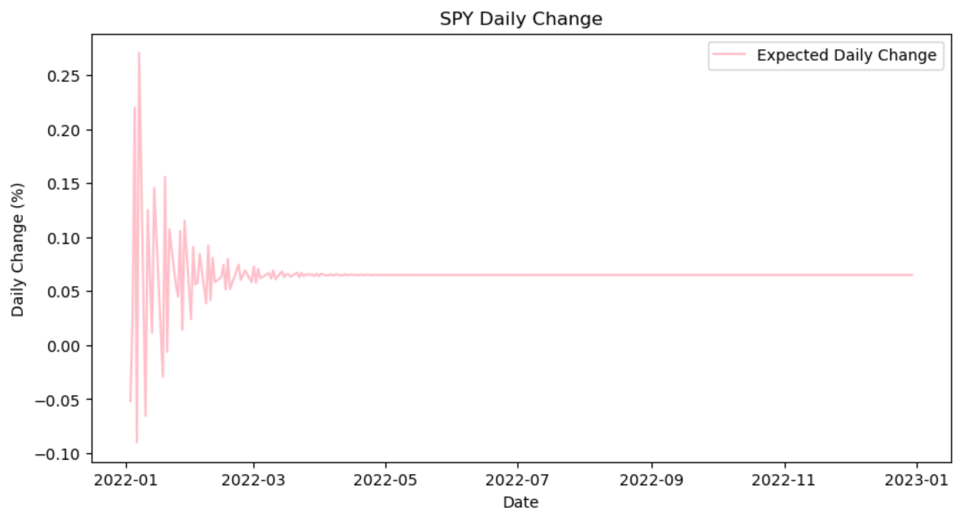

Figure 10. Forecasted Daily Change Values

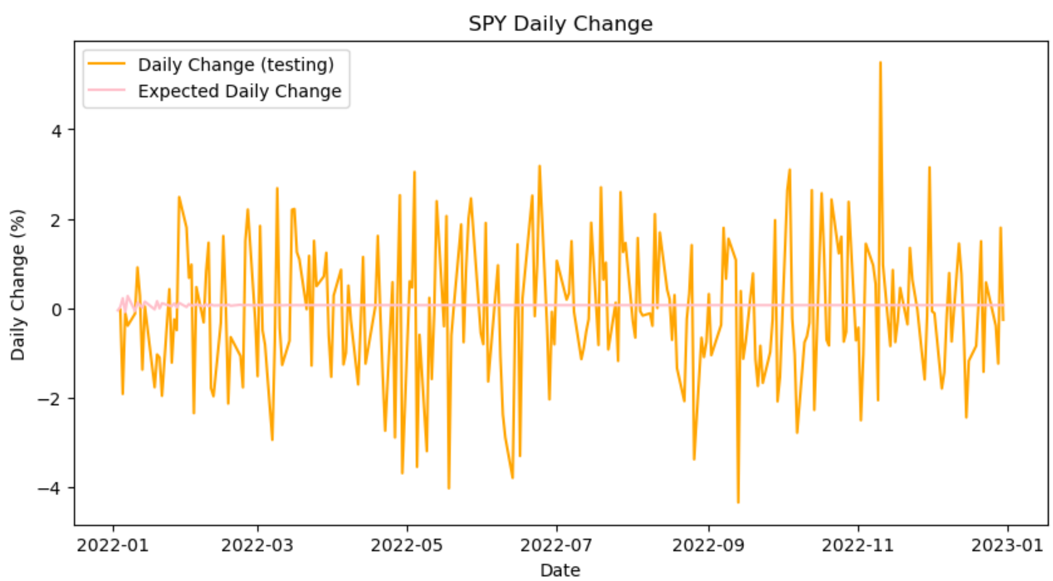

Figure 11. Forecasted Daily Change compared to Actual Daily Change (graph)

	Daily Change	Expected Daily Change	Daily Change (residuals)
1	-0.03	0.03	-0.06
2	-1.92	0.22	-2.14
3	-0.09	-0.09	-0.00
4	-0.39	0.27	-0.67
5	-0.12	-0.07	-0.06
6	0.91	0.12	0.78
7	0.27	0.06	0.21
8	-1.38	0.01	-1.39
9	0.04	0.15	-0.11
10	-1.77	-0.03	-1.74

Figure 12. First 10 Forecasted Daily Change Values compared to Actual Daily Change (table)

3.2 Comparison to other methods

In this section, I'd like to compare the effectiveness of a trading strategy using this ARMA model to other baseline methods. For Model #1, we buy the stock when we forecast a positive change and sell when we forecast a negative change. Here, we lose $32.69. For Model #2, we use a method from Technical Analysis: Simple Moving Averages. In this model, we compute the rolling 20-day and 50-day averages. If the 20-day average surpasses the 50-day average, we buy the stock and when the converse happens we sell. This strategy leads to a loss of $70.99, visualized in figures 12 and 13.

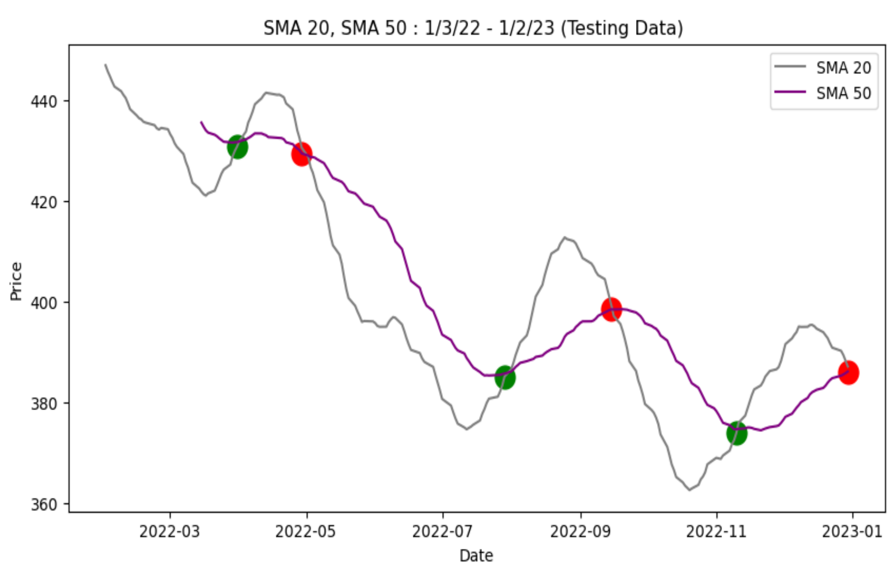

Figure 12. Buy (green) and sell (red) signals

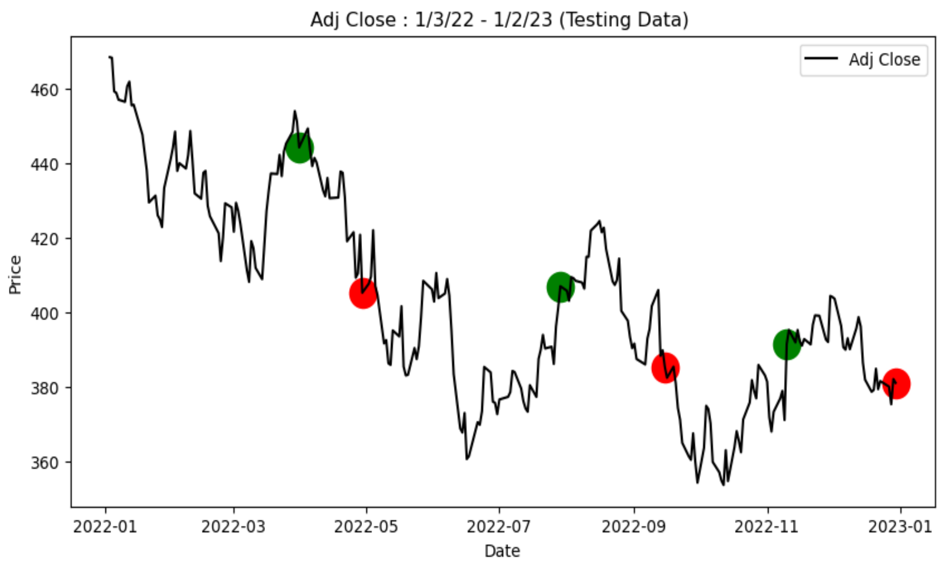

Figure 13. Actual Buy and Sell points

In the last model, we randomly buy and sell 3 times each, per trial. 1000 trials were run and on average, this strategy leads to an average loss of $28.37. This is shown in figure 14 below. For the SPY, during this time-frame random buying has out-performed our ARIMA strategy. Although the ARIMA (6,0,2) model was adequate, using a rolling forecast or a shorter testing period may have led to a more accurate forecast. Furthermore, the forecasts seem to be mean-reverting (according to figure 10).

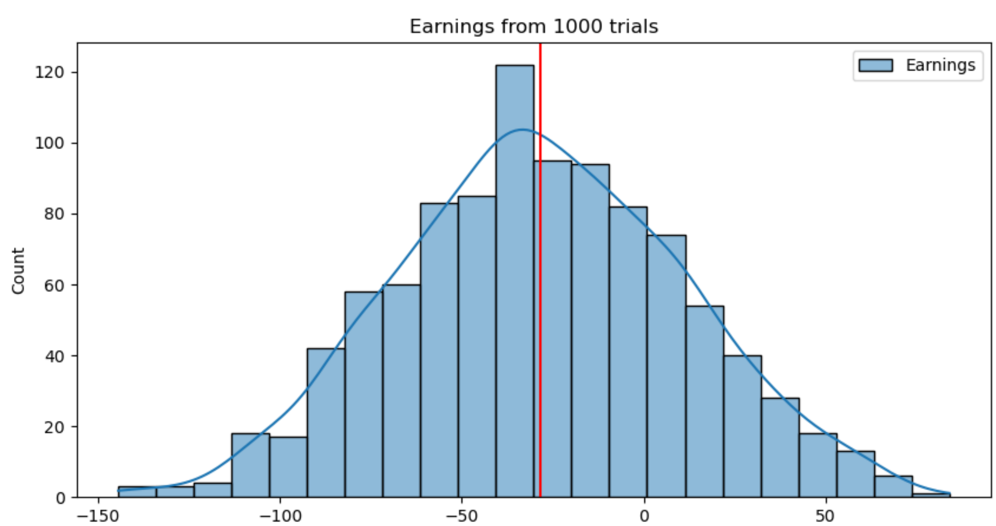

Figure 14. Density plot of 1000 trials of random buying/selling

4. References

[1] Cryer, J. D., & Chan, K. (2011). Time series analysis with applications in R. New York: Springer.

[2] NYSE. (2023, May 21). SPDR S&P 500 ETF Trust (SPY). Retrieved May 1, 2023 from https://finance.yahoo.com/quote/SPY/.