John Lagattuta
Camille Postaer
Nathan Diekema
GSB 570
June 5, 2022

Financial Time-Series Project: **Final Report**

SPDR S&P 500 Trust ETF

A. Overview of the Asset & Market

SPDR S&P 500 Trust ETF tracks the Standard and Poor's 500 Index. This index is made up of 500 large-cap US stocks and serves as a benchmark for the US equity market. The SPDR S&P 500 Trust ETF is traded on the NYSE Arca exchange under the symbol SPY. The SPY ETF was the first index ETF, fully replicating and setting its target price at 10% of the S&P 500. This ETF allows investors to own a small fraction of the entire S&P 500 under one asset for a relatively low cost. Since its creation in 1993, SPY ETF has grown from $6.53 million in assets to almost $417 billion in assets making it the largest ETF in the World. Over one-quarter of the SPY ETF is invested in the information technology sector with top companies such as Apple, Microsoft, and Amazon. It is managed by professionals that use large blends of growth and value stocks to ensure that it accurately follows the index. Below is the breakdown of SPY ETF.

Information Technology	28.02%
Healthcare	13.61%
Consumer Discretionary	12.02%
Financials	11.11%
Communication Services	9.36%

B. One Application for the Econometric Analysis

The S&P 500 represents a diverse pool of the top 500 publicly traded companies in the United States. Its portfolio pulls from some of the most important economic sectors nationwide, including information technology, healthcare, and consumer discretionary. In turn, the current valuation and trajectory of the S&P 500 is relatively indicative of the overall state of the stock market. More importantly, it reflects investor confidence which often correlates with the health of the overall economy. As a consequence, its performance is scrutinized by everyone from investors and speculators to economists, researchers, and government agencies.

Since the SPY ETF represents the S&P 500 index, an econometric analysis of its characteristics would be extremely beneficial for investors and hedge fund managers in detecting and predicting patterns in stock market volatility. Economists could use this information to forecast future GDP growth or potential economic depressions. Government agencies could analyze the effects of various fiscal and monetary policies on the associated movements in the SPY ETF. Managers could even use such econometric analyses of the SPY ETF to better understand when a good time to expand or downsize would be. Overall, having a relative comprehension of what the economic future holds can help anyone interpret how certain events will either directly or indirectly affect them.

C. Properties of the Time-Series

I. Descriptive statistics

Figure 1.1A: Sample Statistics for the SPY and S&P 500 Prices

	Daily		*Monthly*	
	SPY	S&P 500	SPY	S&P 500
Min	25.24	431.90	25.72	440.20
1st Quartile	71.25	1022.60	71.57	1020.60
Median	93.96	1293.20	94.33	1292.30
Mean	131.43	1573.80	132.34	1584.00
3rd Quartile	174.68	1995.20	174.61	2003.4
Max	476.23	4796.60	471.83	4766.20
Skewness	1.55	1.41	1.53	1.40
Kurtosis	1.89	1.71	1.80	1.62

Figure 1.1B: Sample Statistics for SPY and S&P 500 Log Prices

	Daily		*Monthly*	
	SPY	S&P 500	SPY	S&P 500
Variance	0.489	0.309	0.493	0.311
Standard Dev	0.699	0.556	0.702	0.558

The mean for SPY monthly prices is $131.43 and the standard deviation of the SPY monthly log prices is 0.699. The corresponding monthly values for the S&P 500 are $1,584.00 and 0.702. As mentioned in the previous report, the beauty of the SPY is that it mimics the returns of the S&P 500 at a discounted investment price. The sample skewness for SPY prices, 1.53, is extremely positive indicating a highly skewed distribution. Similarly the skewness for S&P 500 is extremely positive at 1.40. The sample excess kurtosis values for SPY and S&P 500 are 1.80 and 1.62, respectively, indicating that the tails of the histograms are fatter than the tails of a normal distribution. See Figure 2.2 for a visual of the monthly prices of both time series from 1993 to 2022.

The mean and standard deviation for SPY and S&P 500 daily prices are relatively equivalent to the monthly statistics. Both time series have daily price patterns consistent with their monthly counterpart. Similarly, the excess kurtosis values are relatively equivalent implying that the tails of the daily distribution are also fatter than a normal distribution. See Figure 2.1 for a visual of the daily prices of both time series from 1993 to 2022.

Figure 1.2A: Sample Statistics for the SPY and S&P 500 Simple Returns

	Daily		*Monthly*	
	SPY	S&P 500	SPY	S&P 500
Min	-0.1094	-0.1198	-0.1604	-0.1694
1st Quartile	-0.0043	-0.0045	-0.0152	-0.0171
Median	0.0007	0.0006	0.0139	0.0121
Mean	0.0004	0.0004	0.0088	0.0073
3rd Quartile	0.0059	0.0057	0.0368	0.0342
Max	0.1452	0.1158	0.1336	0.1268
Skewness	-0.064	-0.201	-0.552	-0.628
Kurtosis	11.75	10.73	1.10	1.22

Figure 1.2B: Sample Statistics for SPY and S&P 500 Simple Log Returns

	Daily		*Monthly*	
	SPY	S&P 500	SPY	S&P 500
Variance	0.0001	0.0001	0.001	0.002
Standard Dev	0.012	0.012	0.043	0.043

The mean for SPY monthly returns is 0.009 and the standard deviation for SPY log returns is 0.043. Annualized, these values are approximately 0.101 ($.009 \times 12$) and 0.149 ($.043 \times \sqrt{12}$), respectively. The corresponding monthly and annualized values for S&P 500 are 0.007 and 0.043 and 0.084 and 0.149, respectively. Although closely related, SPY has a slightly higher mean and volatility than the S&P 500. The sample skewness for SPY, -0.55, is slightly negative meaning it has moderate asymmetry. Similarly the skewness for S&P 500 is moderately negative at -0.63. Both statistics are reflected in the longer left tails seen in Figure 2.6. The sample excess kurtosis values for SPY and S&P 500 are 1.10 and 1.22, respectively, indicating that the tails of the histograms are slightly fatter than the tails of a normal distribution. See Figure 2.4 for a visual of the monthly returns of both time series from 1993 to 2022 and see Figure 2.5 for a visualization of what a $1 investment would look like in each during the same period.

The mean and standard deviation for SPY and S&P 500 daily returns are approximately zero. Both time series have minimal negative skewness, meaning they are relatively symmetric on a day-to-day basis. However, the excess kurtosis values are significantly larger implying that the tails of the distribution are much fatter than a normal distribution. See Figure 2.3 for a visual of the daily returns of both time series from 1993 to 2022.

Figure 1.3: Sample Statistics for SPY Trade Volume

	Daily	Monthly
Min	5,200	1,808,000
1st Quartile	8,599,325	181,296,400
Median	60,230,900	1,344,541,500
Mean	84,584,795	1,770,970,158
3rd Quartile	119,411,350	2,536,184,400
Max	871,026,300	11,882,352,200
Standard Dev	94,631,262	1,846,391,431
Skewness	2.13	1.72
Kurtosis	7.00	4.08

The mean and standard deviation for SPY monthly trade volume are 1.77 billion and 1.85 billion, respectively. Annualized, these values are approximately 21.2 billion and 6.41 billion, respectively. The sample skewness for SPY, 1.72, is very positive meaning its distribution has significant skew. The sample excess kurtosis value for SPY is 4.08 indicating that the tails of the SPY distribution are significantly fat. See Figure 2.7 for a visual of the monthly trade volume of SPY.

The mean and standard deviation for SPY daily trade volume are 84 million and 94 million. Again, we see that the SPY has a highly positively skewed distribution. The excess kurtosis SPY value is even larger on a daily basis implying that the tails of its distribution are much fatter than a normal distribution. See Figure 2.8 for a visual of the daily trade volume of SPY from 1993 to 2022.

II. Visualization

Figure 2.1: Daily Prices from 1993 - 2022 of SPY vs SP500

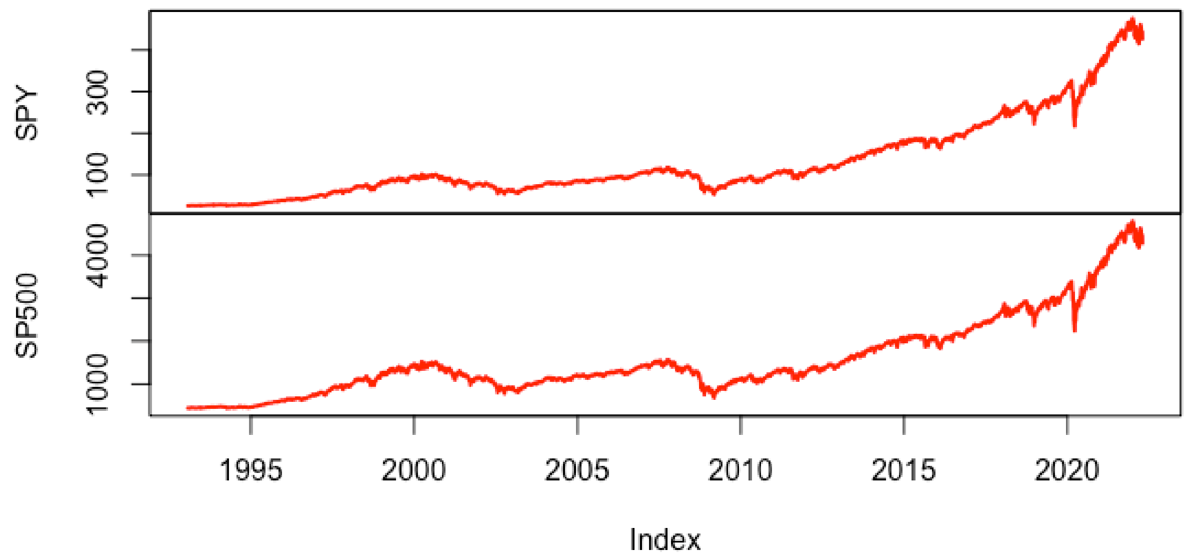


Figure 2.2: Monthly Prices from 1993 - 2022 of SPY vs SP500

Figure 2.3: Daily Returns from 1993 - 2022 of SPY vs SP500

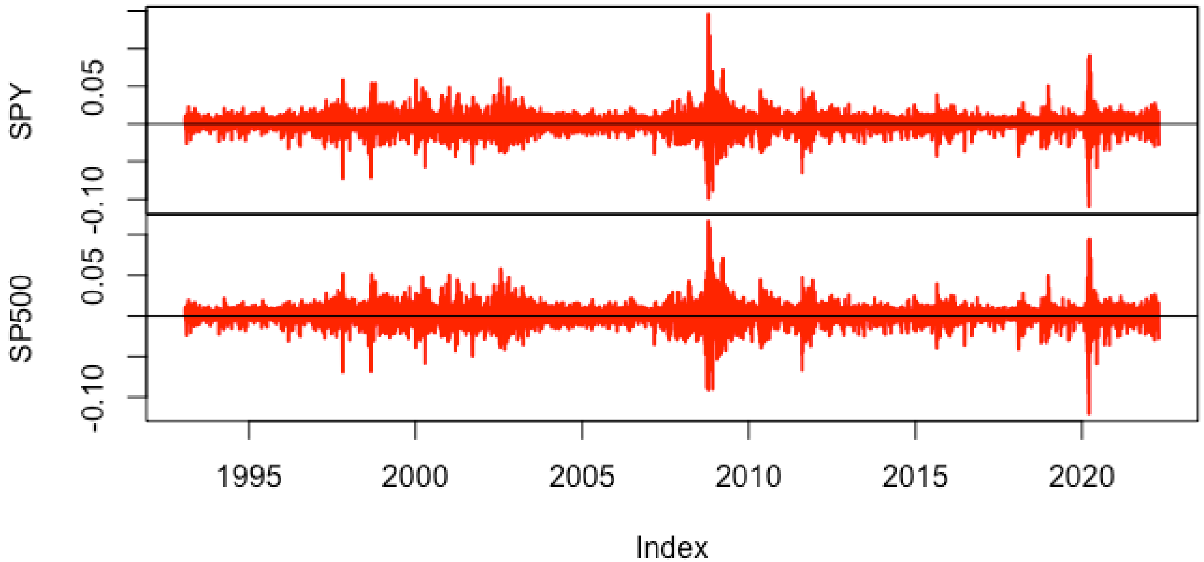


Figure 2.4: Monthly Returns from 1993 - 2022 of SPY vs SP500

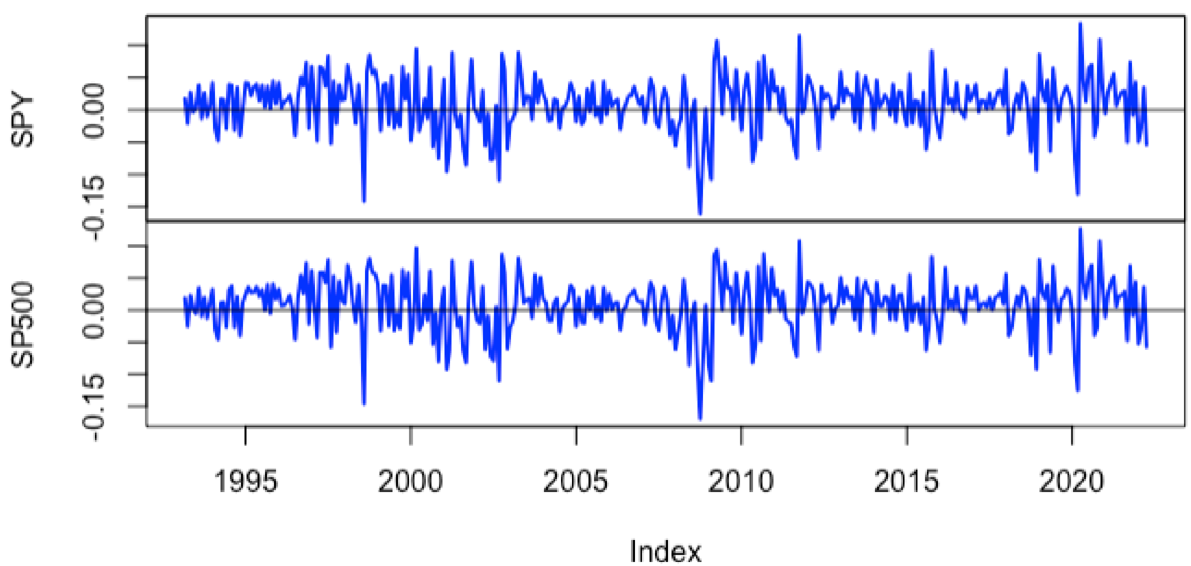


Figure 2.5: Equity curves for SPY and S&P 500 monthly returns

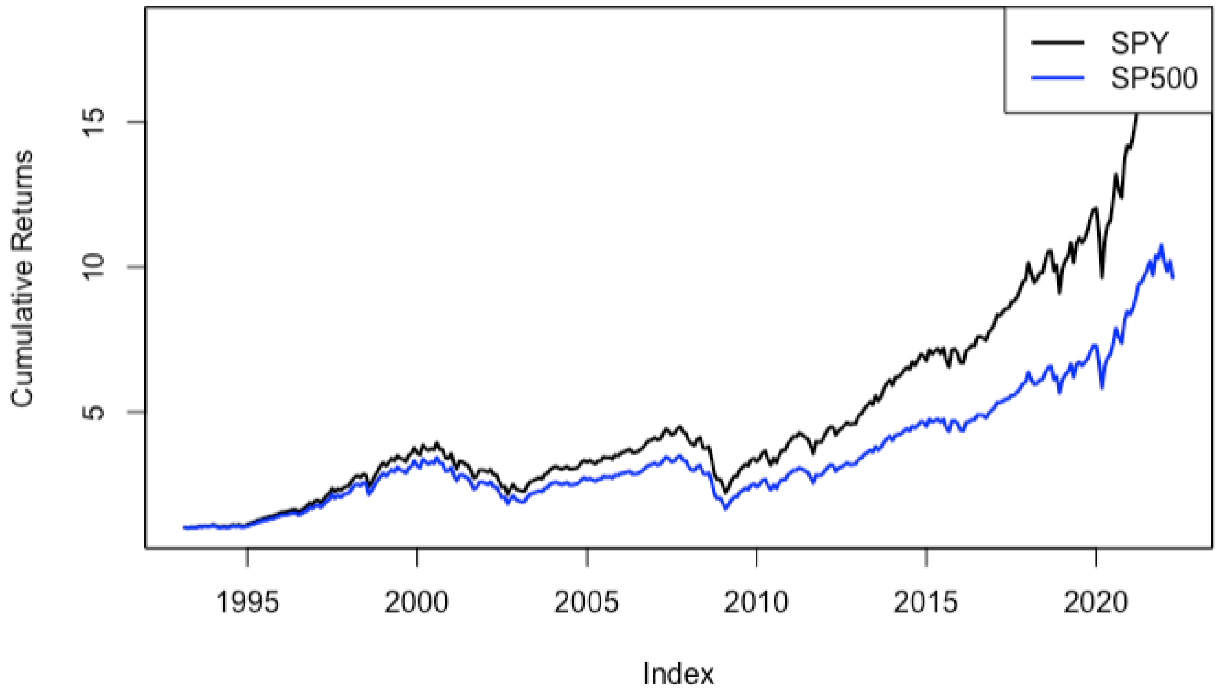


Figure 2.6: Histograms for Daily and Monthly Returns on SPY and SP500 index

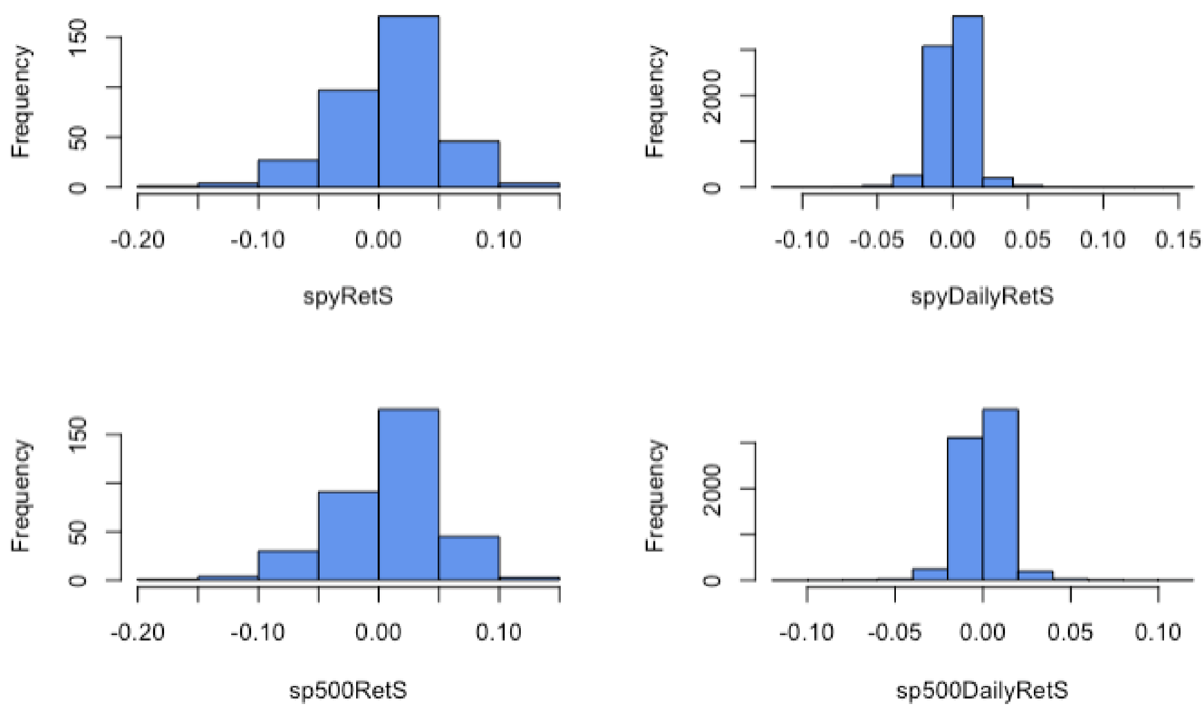


Figure 2.7: Monthly Trade Volume from 1993 - 2022 of SPY

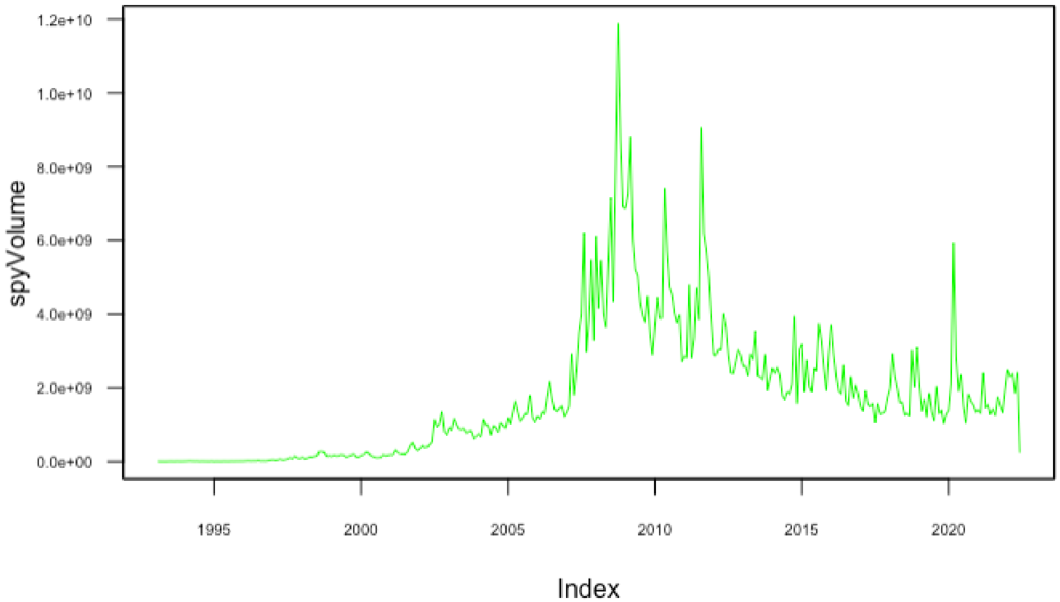
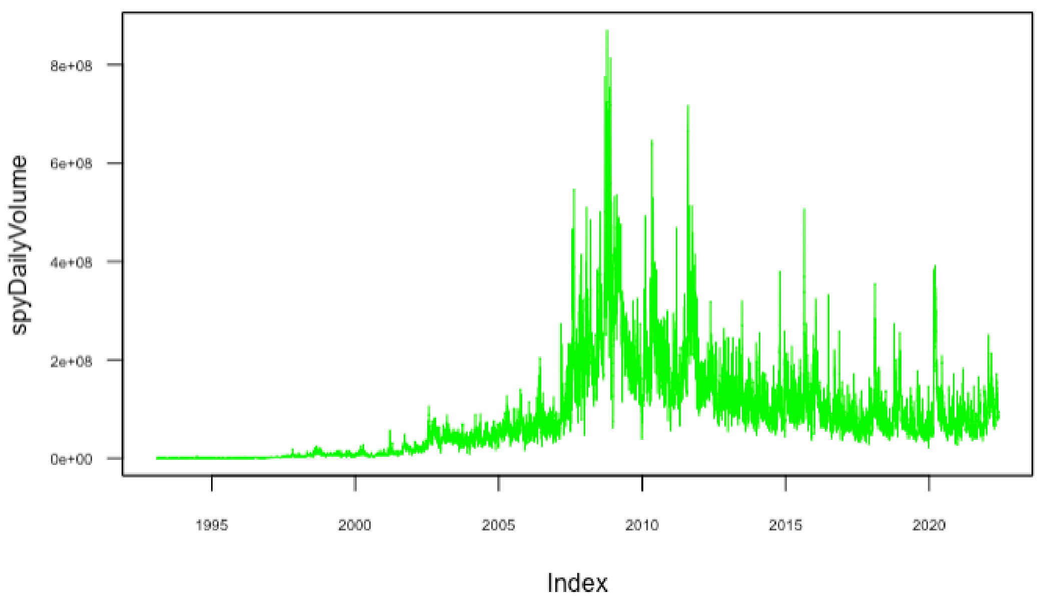


Figure 2.8: Daily Trade Volume from 1993 - 2022 of SPY



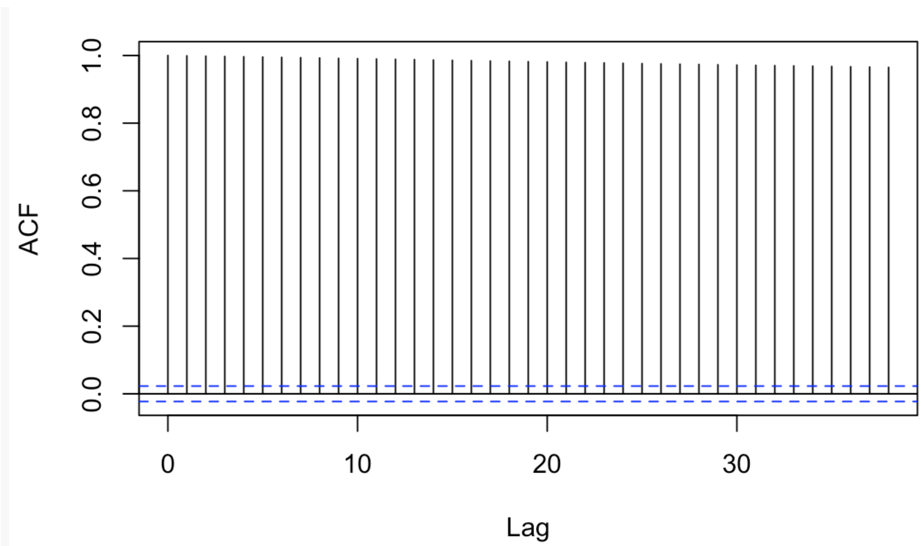
III. Unit-root and seasonality tests

Figure 3.1: Augmented Dickey-Fuller Test

```
	Augmented Dickey-Fuller Test

data:  sp
Dickey-Fuller = 0.34745, Lag order = 19, p-value = 0.99
alternative hypothesis: stationary
```

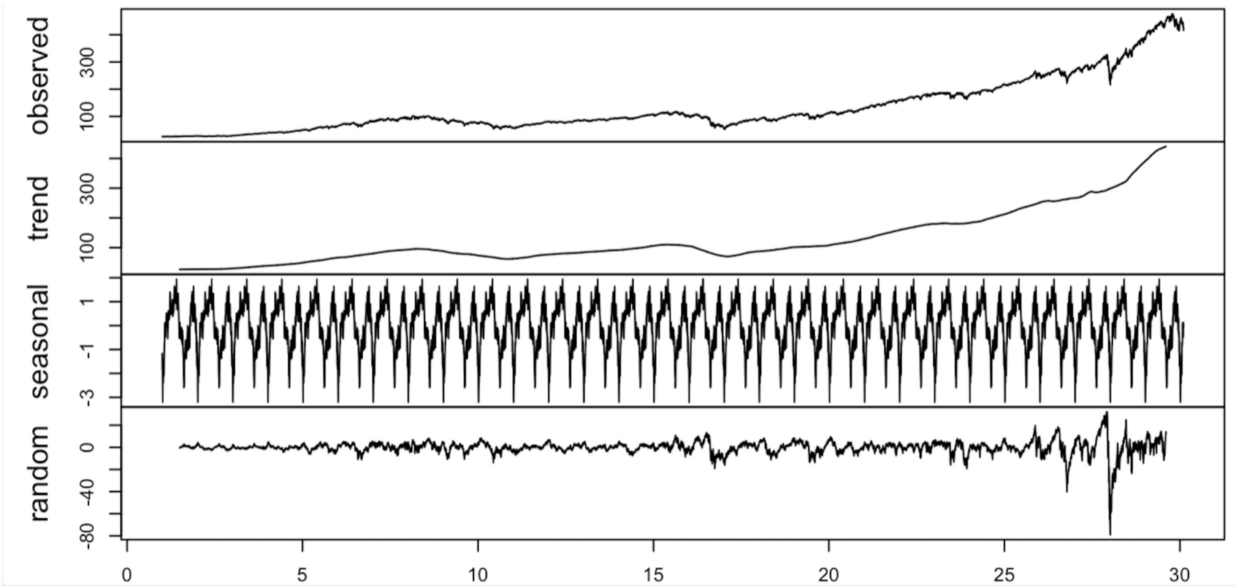
When running a Dickey-Fuller Test on the SPY data, the output displayed an extremely high p-value of 0.99. Because the p-value was so high, the null hypothesis could not be rejected. Unit root was present, meaning that the process was non-stationary and overtime the variance moved towards infinite. This makes sense because for an ETF to be an appealing investment, its price and returns must grow overtime. If the SPY was stationary, no one would invest in it and it would become obsolete.

Figure 3.2: Autocorrelation Function of SPY



Further proof of the nonstationarity of the process can be seen through the autocorrelation function. There is no decay through the lags. The autocorrelation of today's value and any previous lag is near 1. The process does not forget any of the previous values. Again, this is reasonable because historically SPY has had a positive growth and has never fluctuated around any single mean value.

Figure 3.4: Decomposition Function Output



. The output of the decomposition function can be shown in figure 3.4 above. This function breaks down the data by level, first identifying the trend then removing it. Once removed the seasonality can easily be seen. There appears to be seasonality present throughout the day.

D. ARIMA Models of Price and Return

The first step when building an ARIMA model is to ensure the time series is stationary. To make this series stationary a differencing function will be applied. In an ARIMA model, this is done through choosing the optimal value for 'd', or the number of differencing needed to make the series stationary. Typically, the first order of differencing is enough to achieve stationarity. In this section, two ARIMA models will be built: one to model the price of the SPY ETF, and another to model the returns. For the price model, the first order difference is enough to reject the null hypothesis of the augmented dickey-fuller test, indicating stationarity as seen in *Figure 4.1*. For returns there is no need to apply differencing to the series as it is already stationary.

Figure 4.1: Augmented Dickey-Fuller Test on the First Differenced SPY Price

```
        Augmented Dickey-Fuller Test

data:  diff(spy$adjusted)
Dickey-Fuller = -20.234, Lag order = 19, p-value = 0.01
alternative hypothesis: stationary
```

Next, the optimal orders of the Auto Regressive term (p) and Moving Average term (q) must be selected. These refer to the number of lags to be used as predictors and the number of lagged forecast errors that should go into the ARIMA model, respectively. These values can be roughly determined from the PACF and ACF plots. Based on the PACF plots found below, the optimal AR term is likely somewhere between 1 and 8 for price, and around 1 for returns. Similarly, based on the ACF plots, the best value for the MA term is most likely a 0 or 1 for both price and returns. Next, to find the optimal ARIMA model a variety of p, q, and differencing terms will be tested for both price and returns. A total of four models were fitted to the SPY data. The terms for the models were determined by examining the PACF and ACF plots as discussed above as well as using the built in auto arima function. The best model will be chosen based on trained error metrics and AIC, while also favoring simple models.

Figure 4.2: PACF of First Differenced SPY Price

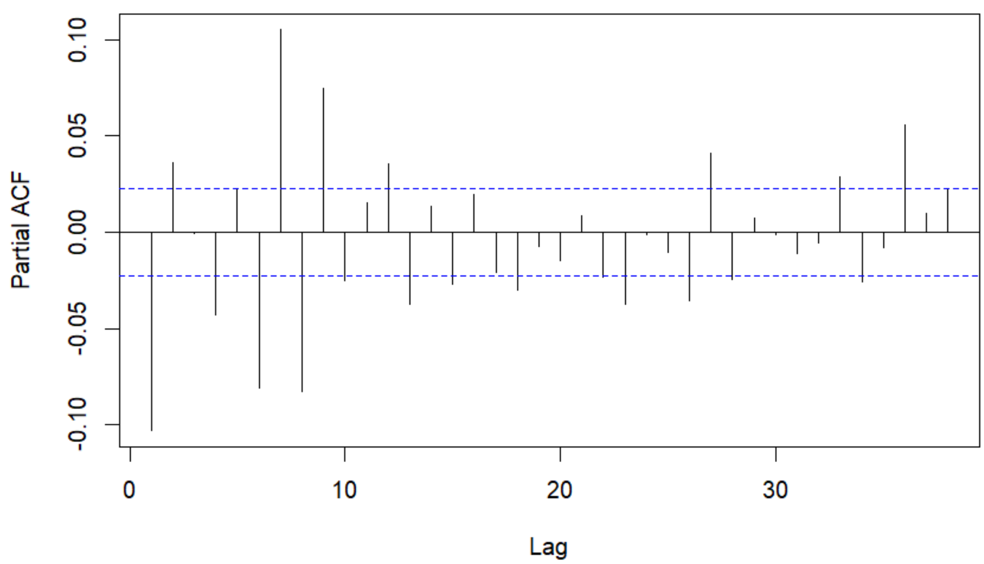


Figure 4.3: ACF of First Differenced SPY Price

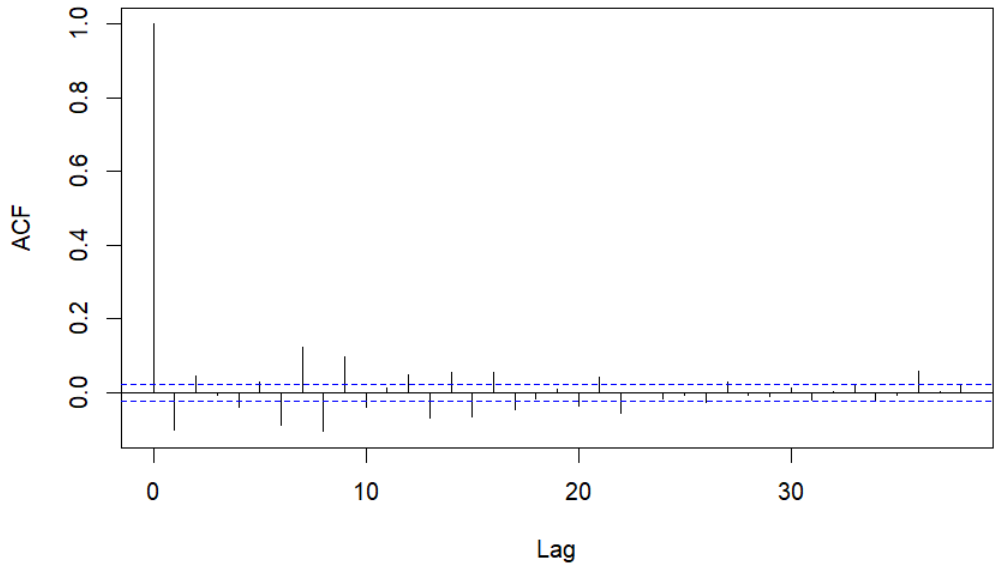


Figure 4.4: PACF of SPY (Log) Returns

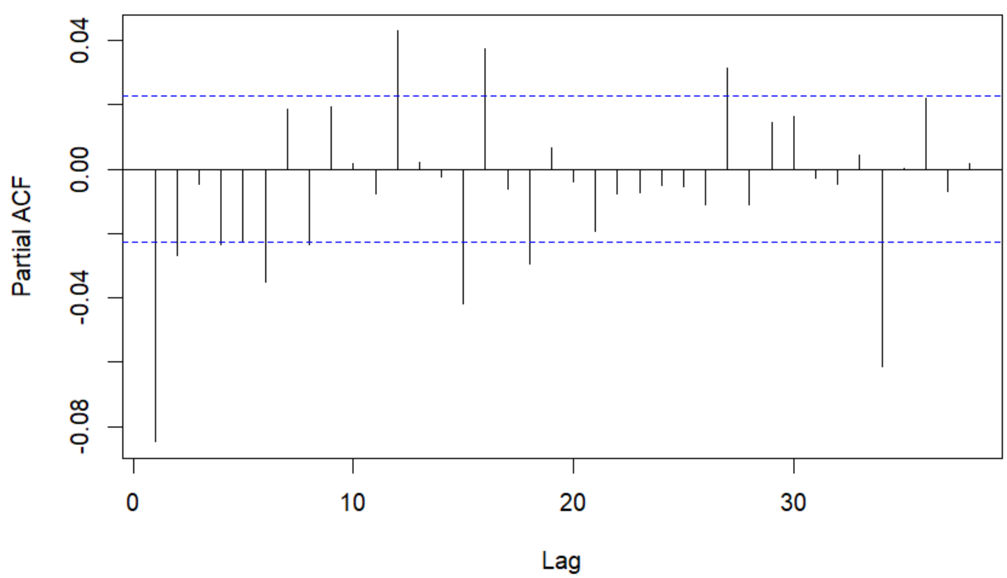


Figure 4.5: ACF of SPY (Log) Returns

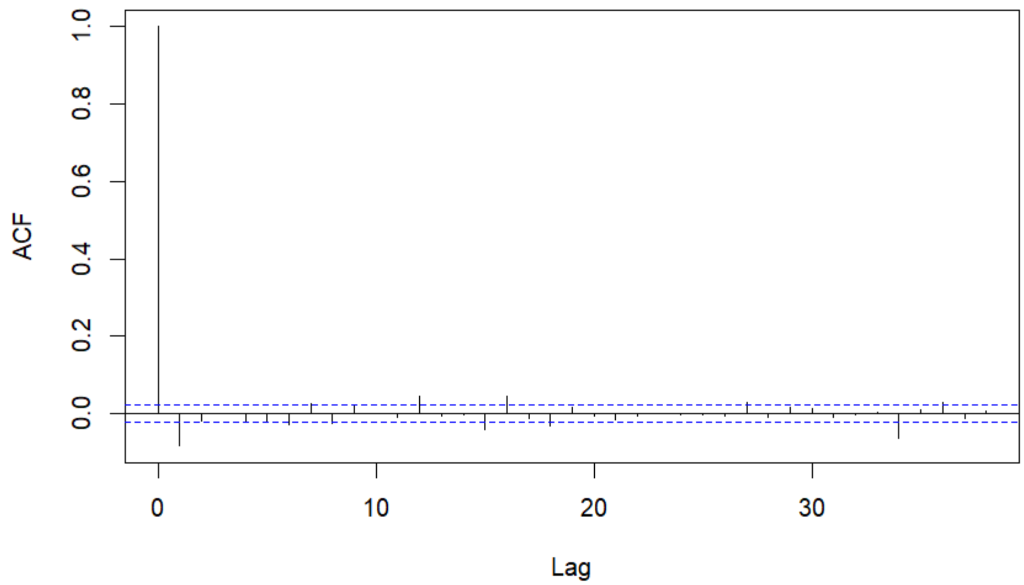


Figure 4.6: Selected ARIMA Model of SPY Price

```
Call:
arima(x = spy_adj, order = c(5, 1, 0))

Coefficients:
          ar1     ar2      ar3      ar4     ar5
      -0.1066  0.0430  -0.0130  -0.0304  0.0157
s.e.   0.0117  0.0117   0.0118   0.0117  0.0117

sigma^2 estimated as 3.417:  log likelihood = -15018.29,  aic = 30048.57

Training set error measures:
                     ME     RMSE     MAE        MPE      MAPE     MASE         ACF1
Training set 0.05747507 1.848431 1.00799 0.03359726 0.7879716 1.001069 0.0003889461
```

Figure 4.7: ARIMA Model Comparison for SPY Price

Model	**AIC**	**Training RMSE**	**Training MAE**
ARIMA(5,2,0)	31344	2.018	1.103
★ARIMA(5,1,0)	30049	1.848	1.008
ARIMA(1,1,1)	30055	1.849	1.007
ARIMA(12,1,1)	29801	1.816	1.014

Figure 4.8: Selected ARIMA Model of SPY Returns

```
Call:
arima(x = rtn, order = c(1, 0, 1))

Coefficients:
         ar1      ma1  intercept
      0.3991  -0.4832      4e-04
s.e.  0.1759   0.1685      1e-04

sigma^2 estimated as 0.0001394:  log likelihood = 22305.63,  aic = -44603.25

Training set error measures:
                       ME       RMSE         MAE MPE MAPE      MASE         ACF1
Training set 8.215857e-06 0.01180861 0.007823891 NaN  Inf 0.6742567 -0.005244438
```

*Figure 4.9: ARIMA Model Comparison for SPY Returns**

Model	AIC	Training RMSE	Training MAE
★ARIMA(1,0,1)	-44603	11.809e-3	7.8239e-3
ARIMA(1,0,0)	-44598	11.814e-3	7.8231e-3
ARIMA(1,0,4)	-44606	11.802e-3	7.8208e-3
ARIMA(2,0,2)	-44604	11.805e-3	7.8210e-3

**Additional significant figures added for comparison purposes*

After running a variety of ARIMA models, the best performer was determined by comparing training set errors and AIC. The top performing models for both price and returns are shown in figures 4.6 and 4.8 respectively. A summarized comparison of each ARIMA model tested for price and returns is provided in Figure 4.7 and 4.9 respectively. The best models were determined based on AIC value, training set error, and simplicity. The detailed output for the selected forecasting model for price is shown in Figure 4.6. It performed better than most of the other models in terms of AIC while also achieving lower training errors. Moreover, it provides a simpler solution than the ARIMA(12,1,0) model which is important when considering overfitting. As a result, this model was chosen as the optimal solution which is corroborated by our out-of-sample testing comparison found in Figure 4.7.

The best ARIMA model found for forecasting returns is pictured in Figure 4.8. This model, which was suggested by the auto arima function, had the second best AIC score in addition to lower or equal training errors. Moreover, it performs better than the other models in the out-of-sample testing found in Figure 4.9 which is why it was selected as the optimal solution. Next, we will plot the residuals and perform an Ljung-Box test on the models to check each of the models validity. As seen below, both of the models' standardized residuals appear stationary which is corroborated by the Autoregressive plots. Moreover, there appears to be no sign of serial correlation in the residuals of the ARIMA model chosen.

Figure 4.10: Checking ARIMA Model of Price

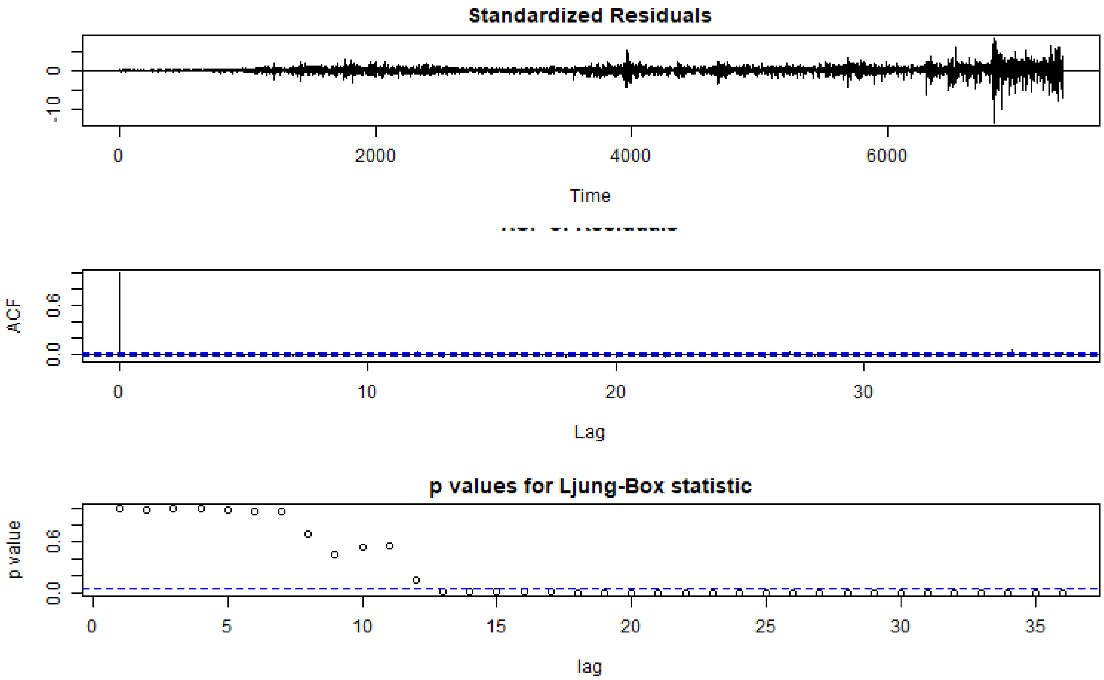
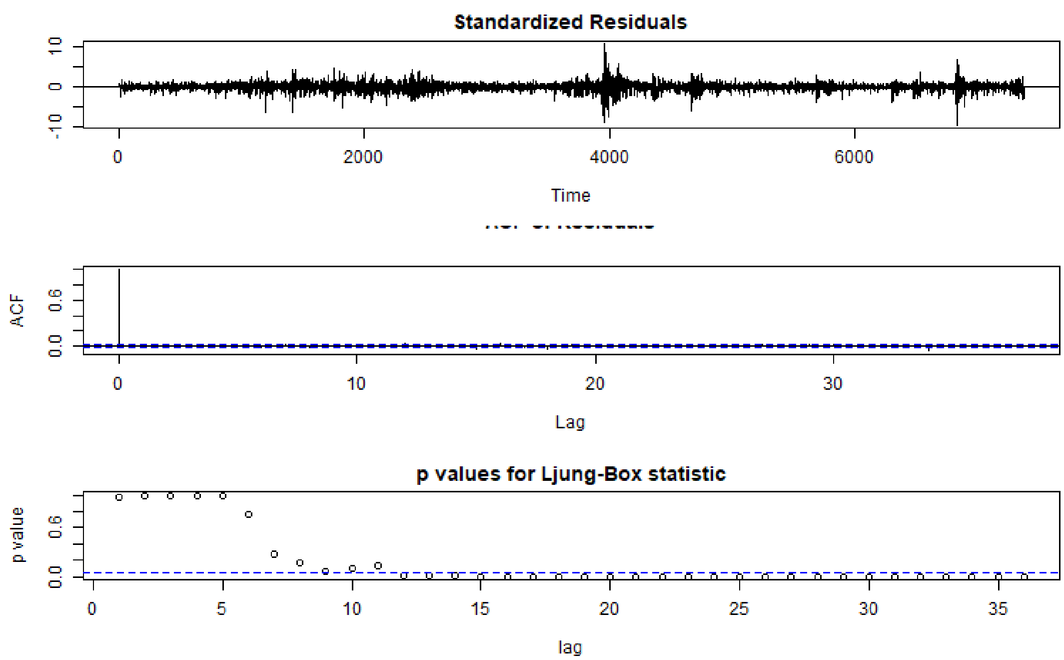


Figure 4.11: Checking ARIMA Model of (Log) Returns



i. Forecasting power of the ARIMA models

Figure 4.12: Fitted ARIMA Model Plotted Against Actual SPY Price

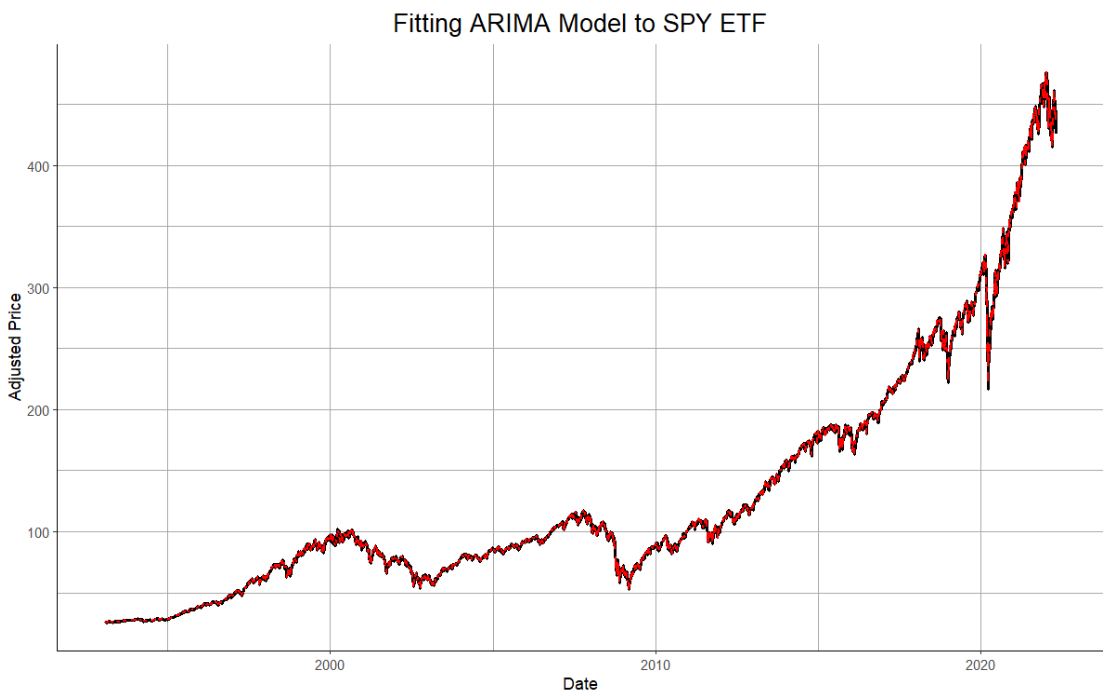


Figure 4.13: Closer look at Fitted ARIMA Model From 2021-2022

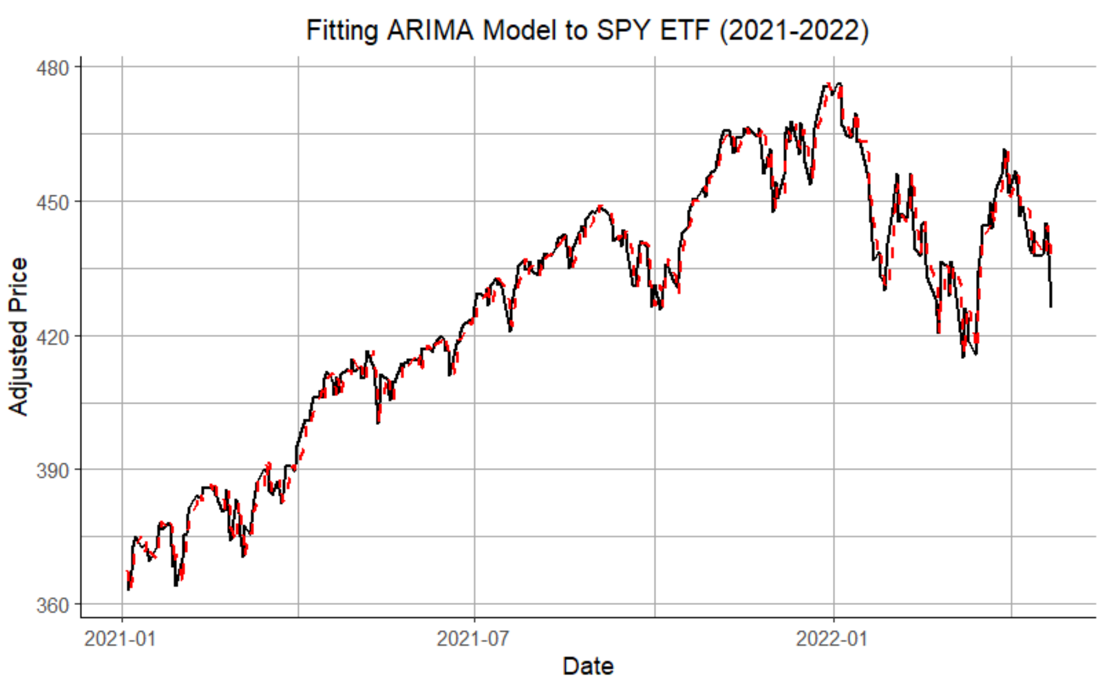


Figure 4.14: Fitted ARIMA Model Plotted Against Actual SPY Returns

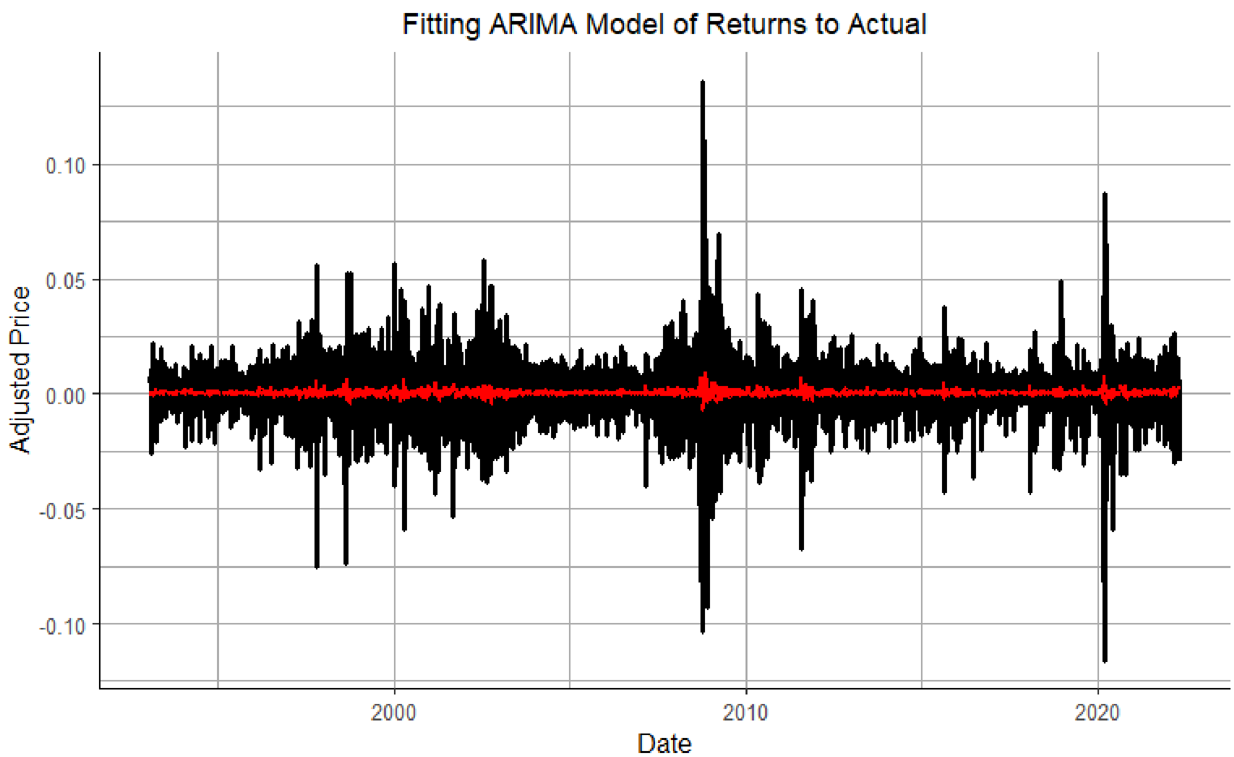


Figure 4.15: ARIMA Forecasted Price 100 Days in the Future

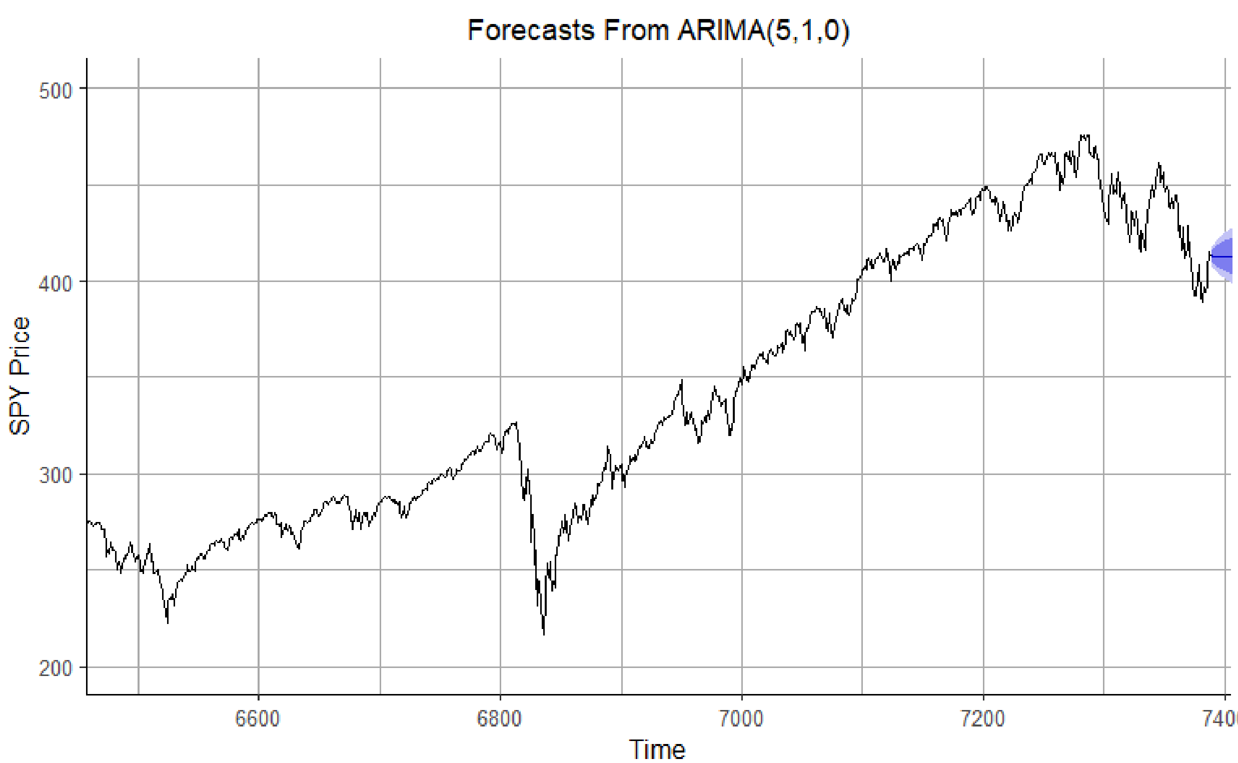
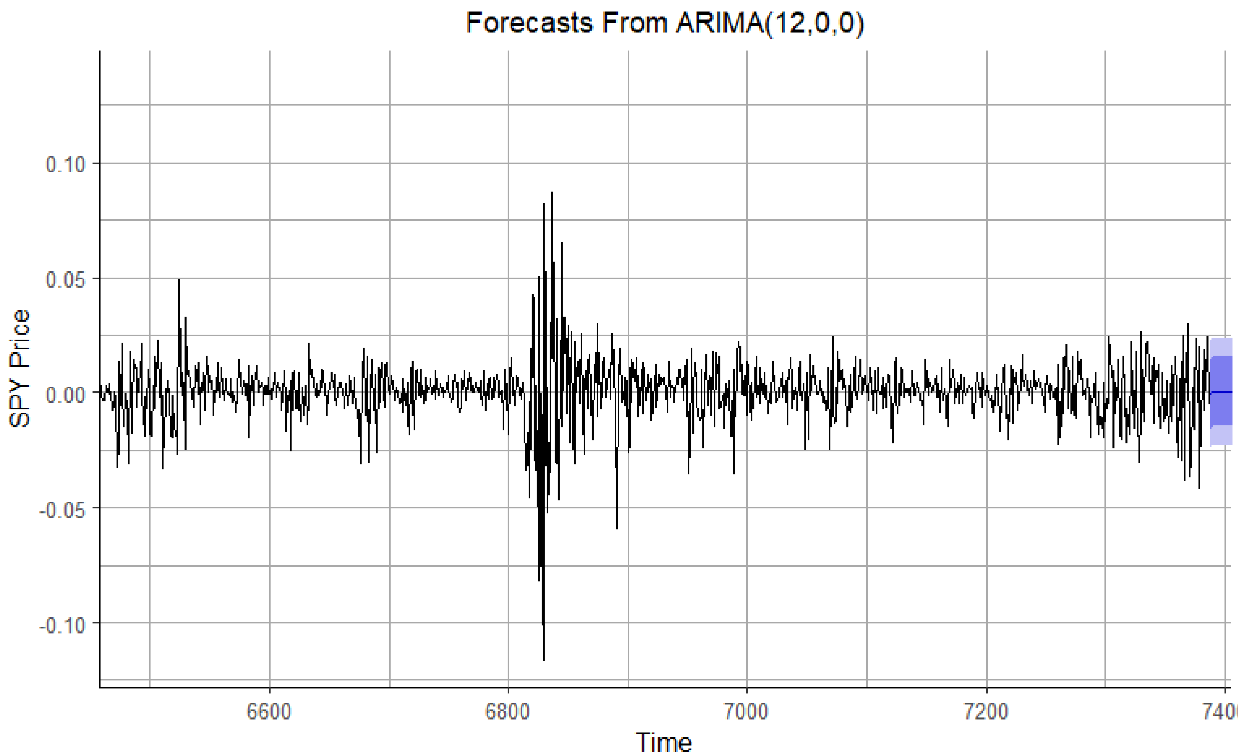


Figure 4.16: ARIMA Forecasted Return 100 Days in the Future



Based on what can be gleaned from the plots above, the chosen ARIMA models appear to be effective forecasters. *Figures 4.12* through *4.14* show the output of the fitted ARIMA models closely following the trends of the actual historical values. After fitting the models to the historical data, both were used to forecast the price and return up to 100 days in the future. The dark blue and light blue shaded regions on the plots represent the 80% and 95% confidence intervals, respectively. The price ARIMA model forecasts that the SPY ETF will gently increase over the next couple months as seen in *Figure 4.16*. The return ARIMA model, on the other hand, forecasts that the returns will fluctuate slightly around the mean (zero) before returning to zero as expected.

To further compare the performance of the models tested, an out-of-sample fit was applied in the form of back testing. The data was divided into a training and testing sample, each consisting of 7,237 and 150 data points respectively. The following results use a rolling of estimation-prediction to compute 1-step ahead forecasts for the given models, starting with forecasts in January, 2022. The results for each model are compared using the root mean squared error (RMSE), and mean absolute error (MAE) obtained through back testing each model. The selected models are marked with a star. It is observed that the models chosen for both the price and returns perform better in terms of both RMSE and MAE than the other models tested.

Figure 4.17: Backtesting Results for All Price Models Tested

Model	RMSE	MAE
ARIMA(5,2,0)	6.796	5.256
★ARIMA(5,1,0)	5.615	4.471
ARIMA(1,1,1)	6.193	4.9438
ARIMA(12,1,1)	6.168	4.877

Figure 4.18: Backtesting Results for All Return Models Tested

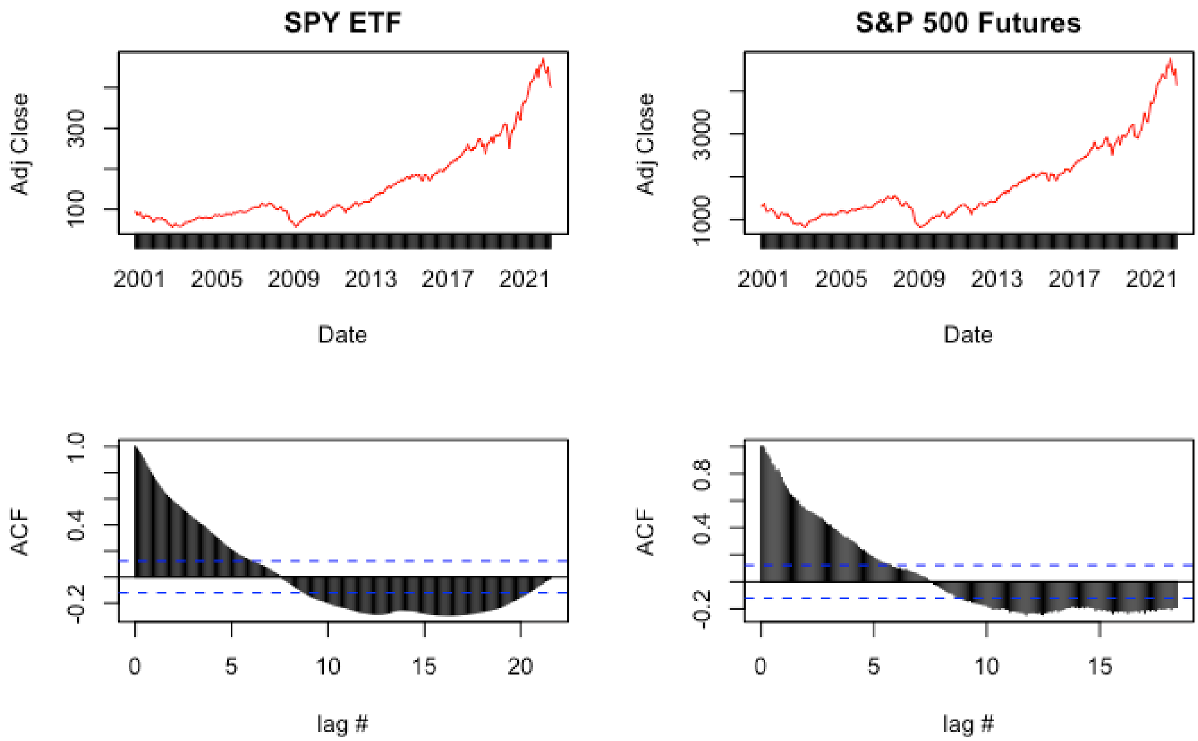
Model	RMSE	MAE
★ARIMA(1,0,1)	1.4338e-2	1.1386e-2
ARIMA(1,0,0)	1.4355e-2	1.1349e-2
ARIMA(1,0,4)	1.4372e-2	1.1382e-2
ARIMA(2,0,2)	1.4339e-2	1.1397e-2

E. Multivariate Analysis Between Spot and Futures Prices

For our multivariate analysis, we will be comparing the SPY closing prices and E-Mini S&P 500 Futures closing prices to determine if they are better suited for a VAR or VECM model. For time series modeling, a vector autoregressive model (VAR) is used to describe short-term dynamics. If the SPY and E-Mini show signs of long-term equilibrium relationships, however, a vector error correction model should be used (VECM). In order to determine this relationship, we must test for cointegration between the spot (SPY) and futures (E-Mini) prices using a Johansen procedure. If the Johansen procedure shows evidence of cointegration, then causality can be tested to examine the discovery role of futures prices.

Before we begin analysis, we should reconfirm that both the SPY and E-Mini closing prices are non-stationary. A visual inspection of the charts below confirms this hypothesis.

Figure 5.1: Adjusted Closing Prices & Corresponding ACF



For statistical evidence, we can use a Box test for stationarity and an Augmented Dickey-Fuller (ADF) test for the presence of a unit root. As seen below, the p-value for the Box-Ljung test was significant for both the SPY and E-Mini, meaning we can reject the null hypothesis of stationarity. The p-value for the ADF test was not significant for both, meaning we fail to reject the null hypothesis and have strong evidence of non-stationarity.

Figure 5.2: Box & ADF Tests for Stationarity or Unit Root Presence

```
        Box-Ljung test

data:  spyPrices
X-squared = 3987.5, df = 25, p-value < 2.2e-16
```

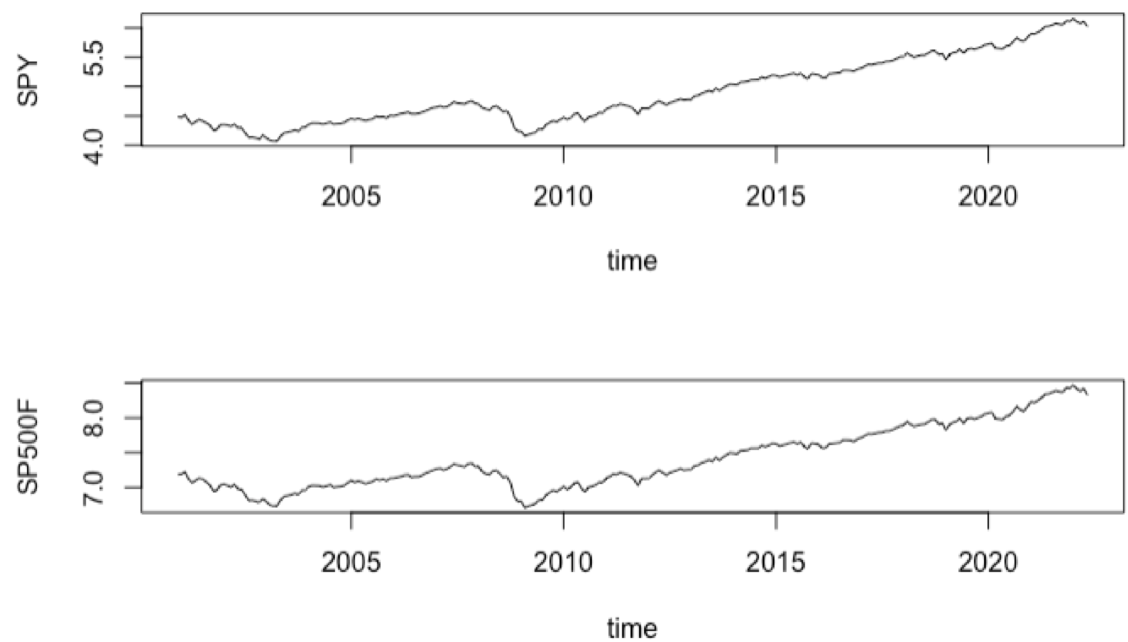
```
        Augmented Dickey-Fuller Test

data:  spyPrices
Dickey-Fuller = -0.82873, Lag order = 6, p-value = 0.9582
alternative hypothesis: stationary
```

```
        Box-Ljung test

data:  sp500futuresPrices
X-squared = 3258.3, df = 25, p-value < 2.2e-16
```

```
        Augmented Dickey-Fuller Test

data:  sp500futuresPrices
Dickey-Fuller = -1.1418, Lag order = 6, p-value = 0.9138
alternative hypothesis: stationary
```

If we take a particular linear combination of these series it can sometimes lead to a stationary series. Such a pair of series would then be termed *cointegrated*. Below are the log prices of the SPY and E-Mini S&P 500 Futures closing prices. Notice how individually the time series are non-stationary, but the difference between them will likely end up being stationary.

Figure 5.3: Log Closing Prices



The basis behind VAR models is that each of the time series in the system influences each other. That is, we can predict the series with past values of itself along with other series in the system. Using Granger's causality test it is possible to test this relationship between the SPY and E-Mini before even building a model.

Figure 5.4: Granger Test for Causality

```
Granger causality test

Model 1: SP500F ~ Lags(SP500F, 1:3) + Lags(SPY, 1:3)
Model 2: SP500F ~ Lags(SP500F, 1:3)
  Res.Df Df      F  Pr(>F)
1    211
2    214 -3 3.5025 0.01634 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The null hypothesis is that the coefficients of past values in the regression equation is zero. In other words, the null hypothesis is that the past values of the SPY do not cause the E-Mini S&P 500 Futures. Since the p-value obtained from the test is less than the significance level of 0.05 ($0.01634 < 0.05$), then we can safely reject the null hypothesis. Therefore, the SPY spot prices cause the E-Mini S&P 500 Futures prices.

The order of integration is the number of differencing required to make a non-stationary time series stationary. In our case, this represents the number of differencing required to make the SPY and E-Mini stationary respectively. If there exists a linear combination of them that has an order of integration less than that of the individual series, then the collection of series is said to be *cointegrated*. When two or more time series are cointegrated it means they have a long run, statistically significant relationship and we should proceed with a VECM model. If not, we proceed with a VAR model.

The first hypothesis of the Johansen procedure, $r = 0$, tests for the presence of cointegration. It is clear in Figure 5.5 that the test statistic does not exceed any of the significance levels ($10.79 < 12.91 < 14.90 < 19.19$), so we have strong evidence to *not reject* the null hypothesis of no cointegration.

Figure 5.5: Johansen Test for Cointegration

```
######################
# Johansen-Procedure #
######################

Test type: maximal eigenvalue statistic (lambda max) , with linear trend

Eigenvalues (lambda):
[1] 0.048056337 0.001310432

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 1 |  0.29  6.50  8.18 11.65
r = 0  | 10.79 12.91 14.90 19.19

Eigenvectors, normalised to first column:
(These are the cointegration relations)

             SPY.l2 SP500F.l2
SPY.l2     1.000000   1.00000
SP500F.l2 -1.231384  -2.31716

Weights W:
(This is the loading matrix)

            SPY.l2     SP500F.l2
SPY.d    0.1713193 -0.0007395309
SP500F.d 0.1676850 -0.0010375646
```

Since r = 0, we can assume there is no statistically significant long run relationship between the SPY and E-Mini prices and we will estimate VAR in differences to periods. To compute various information criteria for our vector time series after a single differencing, we can use the VARorder() function.

Figure 5.6: VAR Model Order

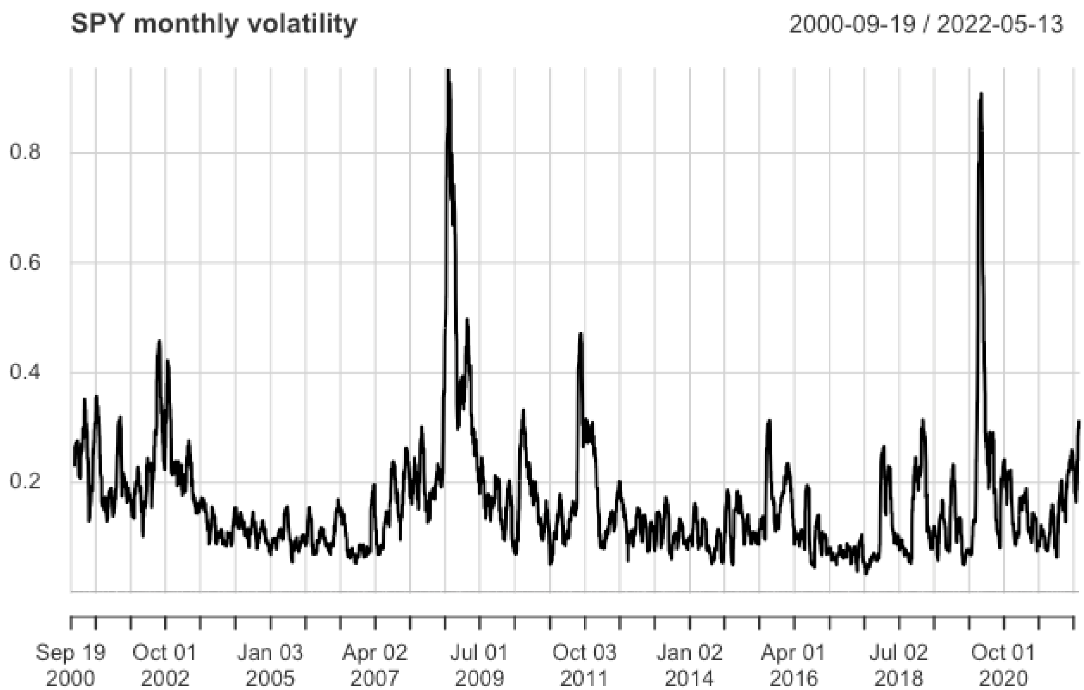
```
d=diffM(xd)
VARorder(d)
```

```
selected order: aic =  7
selected order: bic =  2
selected order: hq =  2
Summary table:
       p      AIC      BIC       HQ    M(p) p-value
 [1,]  0 -16.9201 -16.9201 -16.9201  0.0000  0.0000
 [2,]  1 -17.3385 -17.2768 -17.3136 92.5447  0.0000
 [3,]  2 -17.6547 -17.5313 -17.6049 71.0543  0.0000
 [4,]  3 -17.6512 -17.4661 -17.5765  6.5527  0.1615
 [5,]  4 -17.6952 -17.4484 -17.5956 15.8746  0.0032
 [6,]  5 -17.7179 -17.4094 -17.5933 11.5440  0.0211
 [7,]  6 -17.6882 -17.3180 -17.5387  1.2939  0.8624
 [8,]  7 -17.7226 -17.2907 -17.5482 13.5451  0.0089
 [9,]  8 -17.7101 -17.2165 -17.5108  4.5196  0.3402
[10,]  9 -17.6913 -17.1360 -17.4670  3.2884  0.5108
[11,] 10 -17.6697 -17.0527 -17.4206  2.7508  0.6004
[12,] 11 -17.6752 -16.9965 -17.4012  7.6835  0.1039
[13,] 12 -17.6488 -16.9084 -17.3498  1.7994  0.7726
[14,] 13 -17.6427 -16.8405 -17.3187  5.4254  0.2464
```

F. Conditional Variance Analysis: Various types of GARCH models

Before establishing GARCH models, we should first analyze the volatility of annualized SPY daily returns. We can see in Figure 6.1, there are months with very high volatility and months with very low volatility, suggesting a stochastic model for conditional volatility.

Figure 6.1: Annualized Daily Volatility of SPY Returns



Now we can proceed to test and compare GARCH models. We start with the standard GARCH model where we consider the conditional error term to be a normal distribution and compare its performance to that of an IGARCH, GARCH-M, EGARCH, and GJR-GARCH model. The models were compared using the AIC and BIC metrics in the following table.

Figure 6.2: GARCH Model Comparison

Model	**AIC**	**BIC**
GARCH(1,1)	-6.503	-6.499
IGARCH(1,1)	-6.501	-6.497
GARCH(1,1)-M	-6.562	-6.555
EGARCH(1,1)	-6.543	-6.538
GJR-GARCH(1,1)	-6.535	-6.530

Figure 6.3: Optimal Parameters of standard EGARCH Model

```
Optimal Parameters
------------------------------------
        Estimate  Std. Error   t value Pr(>|t|)
mu       0.00037    0.000078    4.7567    2e-06
omega   -0.26892    0.003138  -85.7019    0e+00
alpha1  -0.14032    0.006319  -22.2063    0e+00
beta1    0.97067    0.000403 2410.0724    0e+00
gamma1   0.16598    0.005929   27.9921    0e+00
```

From the model comparison, the GARCH-M and EGARCH model performed the best according to the AIC and BIC comparison metrics. Between the two models, the EGARCH was selected due to its ability to effectively model asymmetric volatility which is very relevant when modeling trends in the stock market. Figure 6.3 shows the optimal estimated coefficients and their corresponding significance. It is observed that every parameter is significant making this a full EGARCH model. The leverage parameter (alpha1) is significant and shows the effect of the sign at a_{t-1} is negative. The gamma coefficient is also significant with an estimate of 0.17, emphasizing the magnitude of asymmetric volatility in the data. The response to negative shock would be -0.305 whereas the response to positive shock would be 0.025.

Figure 6.4: Ljung Box Test of Standard EGARCH Model

```
Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                      5.827 0.01579
Lag[2*(p+q)+(p+q)-1][2]     6.202 0.01943
Lag[4*(p+q)+(p+q)-1][5]     9.491 0.01243
d.o.f=0
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                      1.915 0.16636
Lag[2*(p+q)+(p+q)-1][5]     6.582 0.06541
Lag[4*(p+q)+(p+q)-1][9]     8.193 0.11812
d.o.f=2
```

Figure 6.4 reports the p-values for various lags of a weighted Ljung-Box test on both standardized residuals and squared residuals. The null hypothesis in each test is that there is no serial correlation between the error terms. Since the p-value is significant across all lags for the standardized residuals, we can reject the null hypothesis and say there is strong evidence of serial correlation. The squared residuals, however, have insignificant p-values across all lags meaning there is not enough evidence to reject the null hypothesis, indicating a lack of serial correlation.

Figure 6.5: Adjusted Pearson Goodness-of-Fit Test for EGARCH Model

Adjusted Pearson Goodness-of-Fit Test:

	group	statistic	p-value(g-1)
1	20	199.3	4.728e-32
2	30	241.2	2.611e-35
3	40	261.5	9.853e-35
4	50	296.1	4.769e-37

Figure 6.5 reports test statistics concerning the goodness of fit of the error. It is used to check if the error term follows the normal distribution. The null hypothesis is that the conditional error term follows a normal distribution. As we can see, the normal distribution is vastly rejected as all the p-values are essentially zero. Since the residuals do not fit the normal distribution we can consider them to be more skewed and assume they follow a student distribution. We can analyze this hypothesis simply by changing the model distribution to be “sstd”.

Figure 6.6: SSTD GARCH Model Analysis

Information Criteria

Akaike	-6.6011
Bayes	-6.5946
Shibata	-6.6011
Hannan-Quinn	-6.5989

Adjusted Pearson Goodness-of-Fit Test:

	group	statistic	p-value(g-1)
1	20	43.50	0.0011072
2	30	63.74	0.0002071
3	40	66.53	0.0039018
4	50	82.35	0.0020058

Figure 6.6 reveals that the SSTD distribution was a slightly better fit of the residuals proven by lower information criteria. Unfortunately, the Pearson test still rejects the null hypothesis that the error terms follow a SSTD. Since the residuals more closely follow a skewed distribution with fat tails, we can test to see how a generalized hyperbolic distribution would fit. We can analyze this hypothesis simply by changing the model distribution parameter to be “ghyp”.

Figure 6.7: SGED GARCH Model Analysis

```
Information Criteria
------------------------------------

Akaike       -6.6032
Bayes        -6.5957
Shibata      -6.6032
Hannan-Quinn -6.6006
```

```
Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     30.08      0.05077
2    30     46.14      0.02275
3    40     54.10      0.05456
4    50     72.73      0.01547
```

Figure 6.7 reveals that the GHYP distribution was a far better fit in terms of the Adjusted Pearson Goodness-of-Fit Test. The p-value is higher than 0.05 for the group size of 20 meaning that the GHYP is a good fit for the error term. Additionally, the information criteria has improved vastly from the base model tested originally.

Figure 6.8: Two Conditional SDs of fitted EGARCH Model Compared with Returns

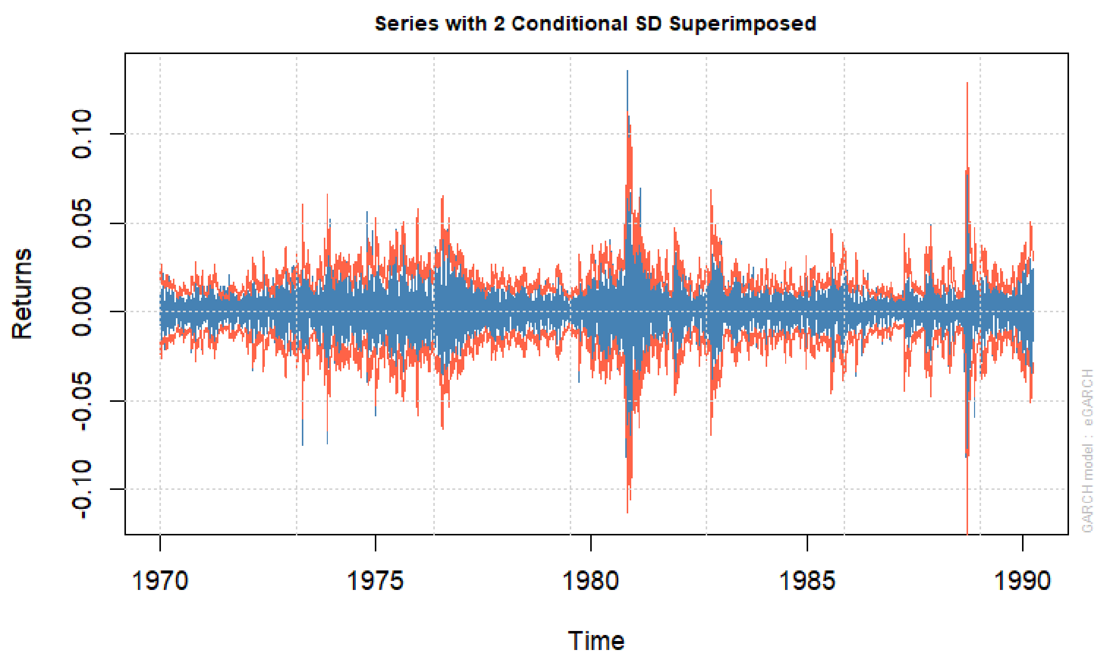


Figure 6.9: QQ-Plot of Standardized Residuals from fitted EGARCH Model

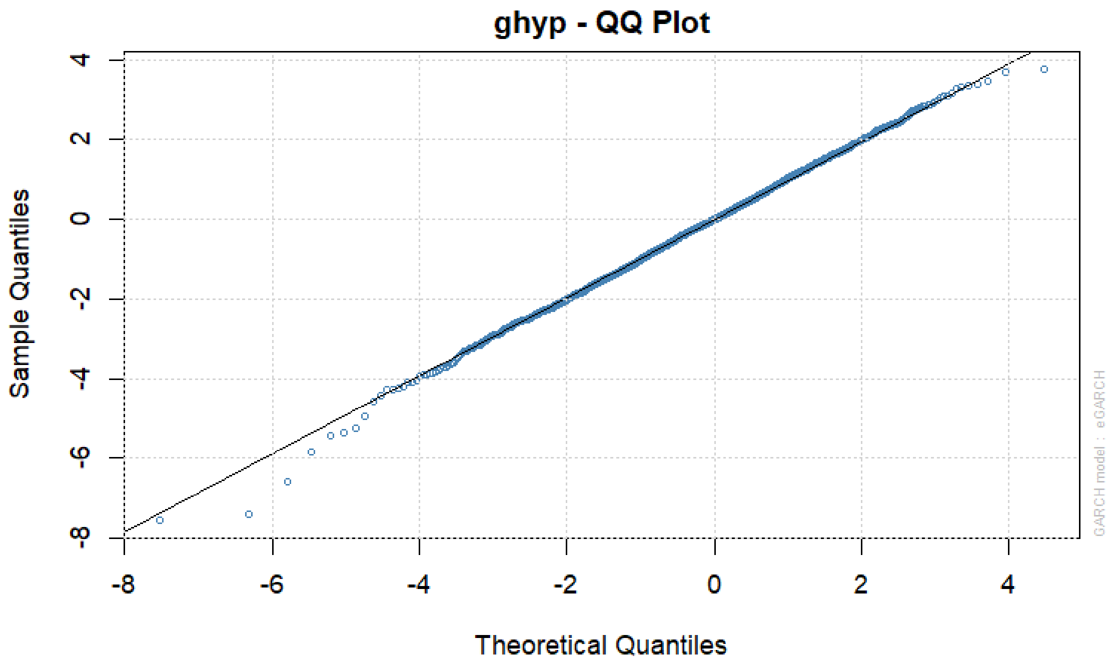
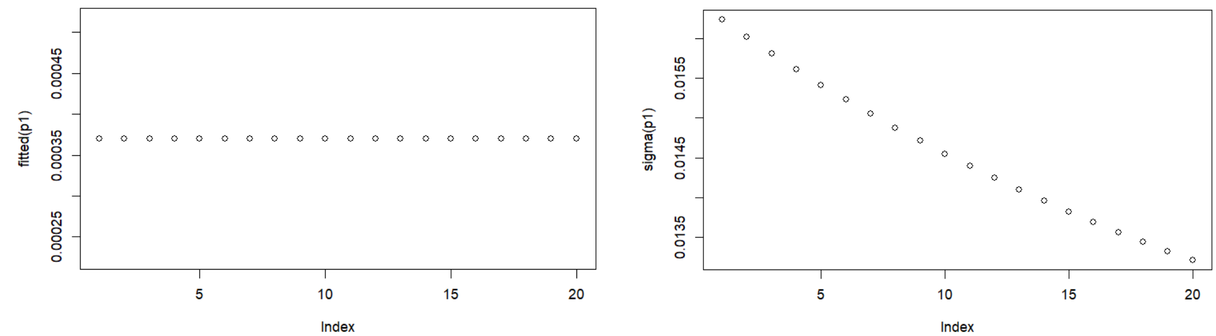


Figure 6.9 illustrates the fit of the fitted volatility model by superimposing 2 standard deviations on the returns of the SPY index. Figure 6.9 shows the QQ-plot of the standardized residuals of the EGARCH model. As we discovered earlier, the distribution of returns is not normal. However, after testing a couple other distributions we were able to achieve a better fitting model using the GHYP distribution. From the QQ-plot, it's clear that the GHYP distribution provides a good fit for the data, effectively accounting for the fat tails and heavy left skew. We can use this model specification to forecast the volatility of SPY daily returns over the next 20 days.

Figure 6.10: GARCH Model Predictions



We can use this model specification to forecast the volatility of SPY daily returns over the next 20 days. The figure above illustrates the volatility and mean predictions on the left and right respectively. Figure 6.10 shows that the model predicts SPY daily return volatility to remain stable over the next 20 days, with increasing uncertainty across the time period. It can also be observed that the mean is slowly approaching the 0.0037, or the mean estimate of the EGARCH model.

G. Value-at-Risk Analysis

This section will discuss the market risk of the SPY ETF through a value-at-risk analysis. There are a couple different methods for characterizing the CDF of loss, but for the purpose of simplicity and transparency the RiskMetrics approach was taken. The RiskMetrics method fits an IGARCH(1,1) model to obtain an estimate of the beta parameter. The coefficient estimates resulting from fitting the SPY ETF is 0.933 as shown below in figure 7.1. The estimate is very close to the standard value of 0.94. The resulting IGARCH model is then used to predict the value at risk (VaR) and expected shortfall (ES) risk metrics. The value of VaR helps determine the extent and respective probabilities of potential losses of the SPY whereas the ES represents the average predicted losses in the worst case scenarios. Both VaR and ES were calculated for four threshold values: 95%, 99%, 99.9%, and 99.99%. The raw results are shown in table 7.1A and the corresponding dollar amounts of VaR and ES given a $1000 position in the SPY are provided in table 7.1B.

Figure 7.1: Fitted IGARCH Model Coefficients

```
Coefficient(s):
       Estimate   Std. Error  t value   Pr(>|t|)
beta 0.93337887  0.00359081  259.935 < 2.22e-16 ***
```

Table 7.1A: Risk Measure Based on RiskMetrics

Probability	VaR	ES
0.950	0.03107	0.03897
0.990	0.04395	0.05035
0.999 *(Risk Management)*	0.05838	0.06361
0.9999 *(Stress Testing)*	0.07026	0.07479

Table 7.1B: VaR and ES at $1000 Position

Threshold	VaR at a $1000 Position	ES at a $1000 Position
0.950	$31.07	$38.97
0.990	$43.95	$50.35
0.999 *(Risk Management)*	$58.38	$63.61
0.9999 *(Stress Testing)*	$70.27	$74.79

Table 7.1C: 10-day VaR and ES at $1000 Position

Threshold	VaR at a $1000 Position	ES at a $1000 Position
0.950	$106.30	$133.48
0.990	$150.63	$172.68
0.999 *(Risk Management)*	$200.33	$218.34
0.9999 *(Stress Testing)*	$241.24	$256.82

From the results obtained through the value-at-risk analysis we can make a few educated predictions about the extent and probability of potential losses. It can be assumed that with a threshold of 95% the greatest expected loss is 3.11% of investment. However, when considering the risk management threshold of 99.9%, the greatest expected loss is 5.84%. In other words, there is a 99.9% chance that the SPY will not drop any more than 5.84% tomorrow. When considering expected shortfall, it can be assumed that within the top 5% of worst case scenarios, the average losses reach approximately 3.8%. The last two tables illustrate the dollar values of VaR and ES given a $1000 position in the ETF. The first table directly corresponds with the 1-day ahead values, whereas the second table provides 10-day ahead estimates. The 10-day ahead forecast indicates that with 95% confidence the maximum losses will only reach $106 over

10 days given an initial investment of $1000. Overall, the SPY fund is a safe investment option with relatively low VaR measures.

H. Conclusion and Managerial Implications

The SPY ETF is one of the oldest and most recognizable US-listed ETFs and is typically a frontrunner in terms of trading volume and assets under management (AUM). The fund tracks the massively popular S&P 500 index, which is representative of a diverse pool of the top 500 publicly traded companies in the United States. Its portfolio pulls from some of the most important economic sectors including information technology, healthcare, and consumer discretionary. In turn, the current valuation and trajectory of the S&P 500, and subsequently the SPY, is indicative of the overall state of the stock market. More importantly, it reflects investor confidence which often correlates with the health of the overall economy. As a consequence, its performance is heavily scrutinized. Since SPY is the most popular ETF that effectively tracks the S&P 500 index, the econometric analyses akin to those covered in this report are extremely beneficial to investors and hedge fund managers in detecting and predicting patterns in stock market volatility.

This analysis began by examining the characteristics of both price and returns, specifically stationarity and seasonality. The augmented dickey fuller test suggests that the future and spot prices are non-stationary processes whereas the log returns are stationary. In terms of seasonality, there is no rule-of-thumb test, but by looking at a decomposition plot by removing trends and other noise the existence of seasonality was clear. There were clear monthly trends with November-April seemingly performing the best whereas May-October had the worst performance. Next, we applied numerous time-series forecasting techniques to model the behavior of the model and to predict what will happen in the future. First, an ARIMA model was fitted for spot price and returns, both of which achieved decent forecasting power. Next, a multivariate analysis was performed comparing the adjusted closing prices of the SPY and the E-Mini S&P 500 Futures to determine if a VAR or VECM model is better suited. From the first hypothesis of the Johansen procedure, r=0, it is clear that there is no statistically significant long-run relationship between the SPY and E-Mini thus VAR is the preferred model. Moreover, the Granger causality test indicates that SPY prices cause E-Mini Future prices. The last time series characteristic examined was volatility. To model the variance of SPY, various GARCH models were assessed, the best of which was an EGARCH(1,1) model with a generalized hyperbolic distribution.

Finally, a value-at-risk (VaR) analysis was performed, where both VaR and ES were calculated for a variety of thresholds. The diverse nature of the SPY ETF makes it a great investment for those looking for a low-risk option. The VaR analysis performed in section G reflects this statement as the maximum expected loss in any given day is relatively low compared to more volatile investments. As a whole, the SPY fund is a safe investment option for investors of all ability levels. This report is meant to be readable and has information that is useful even to the novice investor. After all, having a relative comprehension of what the economic future holds can help anyone interpret how certain events will either directly or indirectly affect them.