A Time Series Approach to Forecasting S&P 500 Returns

Radhika Kulkarni

Abstract

This article addresses the question of predicting S&P 500 daily returns through SPDR S&P 500 ETF (SPY) time series, taken over the 30 year period from 1993-01-29 to 2023-03-31. Using the predictor variables of Consumer Price Index, Unemployment Rate, and CBOE Volatility Index (VIX), we conclude that the ARIMA(1, 0, 1) is better for forecasting SPY adjusted closing price due to its comparatively lower AIC and RMSE values, as opposed to other tested models. However, this model, which contains the variables SPY, UNRATE, CPI and VIX, is more complex and hence less interpretable. We were able to accurately predict the adjusted closing price to a margin within 1-3 dollars from the actual value, with greater deviations as we move farther away from the knowledge base. Impact and significance of our results include improved understanding of which predictor variables can better explain S&P 500 returns, application to improvement of public financial literacy, and models which can be used and updated to predict future returns by interested parties.

Introduction and Objectives

The stock market is a tricky creature. In a free market, buyers invest and sell stocks for their personal and monetary gain, but this fluctuates based on extreme volatility in the market. The most diversified and perhaps most attractive stock exchange to the public is the S&P 500, which is often considered to be more robust to inflation. This inflation has been more pervasive in its influence in the last 10 years, especially taking into account the pandemic stock drop, supply chain issues, semiconductor chip shortage, tech layoffs, and other major events that may have influenced the volatility of the stock market close price. Despite these favorable characteristics of the S&P 500, it is hardly affordable for the average person to buy, manage, or sell all 500 of the largest companies listed on stock exchanges. In our analysis we will use SPY time series, a more accessible type of investment fund that is traded on the stock market; the value of one share of this ETF is approximately 1/10 of the cash value of S&P stock. Given the SPY time series from 1993-01-29 to 2023-03-31, our goal is to understand factors affecting the volatility of the stock market and forecast future stock market daily returns, assuming we invest a certain amount of money today.

In our initial analysis, we consider VIX as a possible predictor time series, since many research papers found it to be an influential factor in determining the daily stock market yield returns, alongside S&P 500. Additionally, we include variables CPI and UNRATE, which are also found in the literature to be useful for prediction.

Data Description

In our study, we will analyze the response variable and our possible predictors in order to fit a model to SPY daily returns. Our goal is to be able to predict, within a margin of error, future SPY daily returns.

Response Variable: SPY

We obtained the SPY dataset from Yahoo Finance, using the available R package “quantmod”. There are six columns in the daily SPY dataset - Date: the date on which the stock market daily index value was recorded at market close; Open: daily index value at market open; High: highest index value of the day; Low: lowest index value of the day; Adj. Close: the adjusted daily index value at market close; and Volume: the number of shares traded on a given day. The stock market closes at 4 pm Eastern Time, except on holidays when the stock market may close earlier. It is the Adjusted Closing column that will be of interest to us. *Going forward, the term “SPY series” refers to the Adjusted Closing SPY value.*

Predictor Variables: CPI, UNRATE, and VIX.

During our research of financial literature, we discovered that, in other studies, certain predictor series are used to predict S&P 500 returns. As a result, we add three covariates: Consumer Price Index (CPI), Unemployment Rate (UNRATE), and Chicago Board Options Exchange's CBOE Volatility Index (VIX) to our dataset.

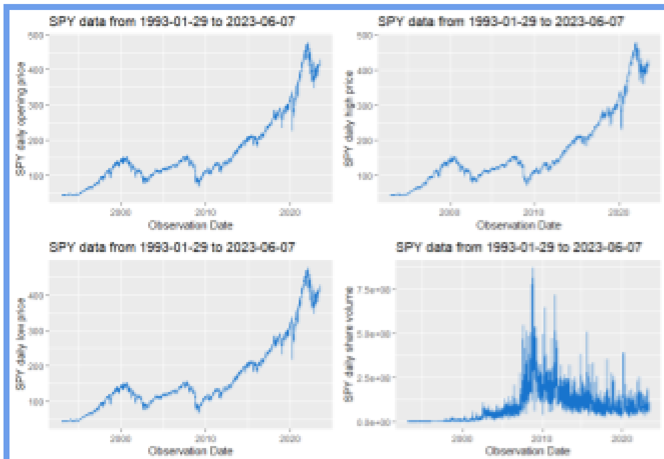


Figure 0 shows the plots for daily opening price, daily high, daily low, and daily share volume. All these plots look similar to the daily adjusted closing plot (shown in Figure 2, top left), except for the daily share volume. For the daily share volume plot, we see a general rising trend between 2005 and 2010, and an overall decreasing trend between 2010 and 2020. Moreover, the daily share volume is less smooth and more volatile than the other variables. This plot follows an inverse pattern compared to the other variables.

1. CPI is a growth rate that quantifies the change in price of consumer-bought goods and services; it is the ratio $\frac{currentvalue - previousvalue}{previousvalue}$.
2. The UNRATE is expressed as a percent, and is the same ratio as CPI, but multiplied by 100. We divide UNRATE by 100 to keep the variable's unit in decimal form.
3. VIX is a ticker name for the market's expectation of the 30 day volatility through the basis of the S&P 500 index options. We align the series' lengths to daily frequency for CPI and UNRATE, which are reported monthly, since VIX and SPY are recorded daily.

Exploratory Data Analysis

In order to uncover the barriers to stationarity, we plot the raw time series for each of SPY, UNRATE, CPI, and VIX variables.

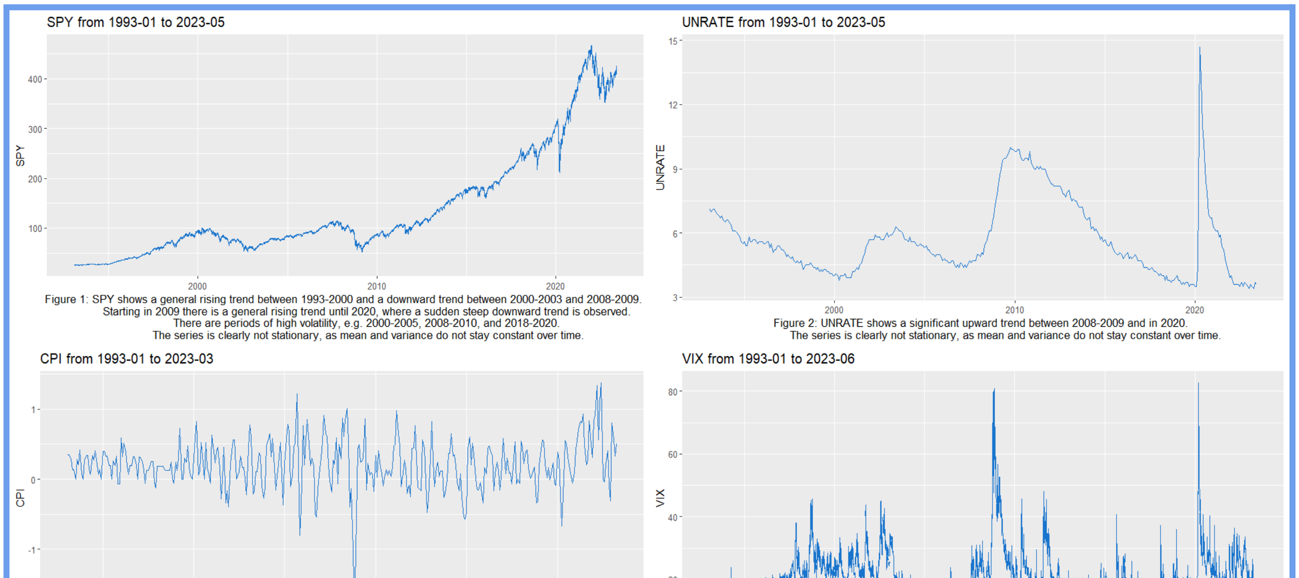


Figure 1: SPY shows a general rising trend between 1993-2000 and a downward trend between 2000-2003 and 2008-2009. Starting in 2009 there is a general rising trend until 2020, where a sudden steep downward trend is observed. There are periods of high volatility, e.g. 2000-2005, 2008-2010, and 2018-2020. The series is clearly not stationary, as mean and variance do not stay constant over time.

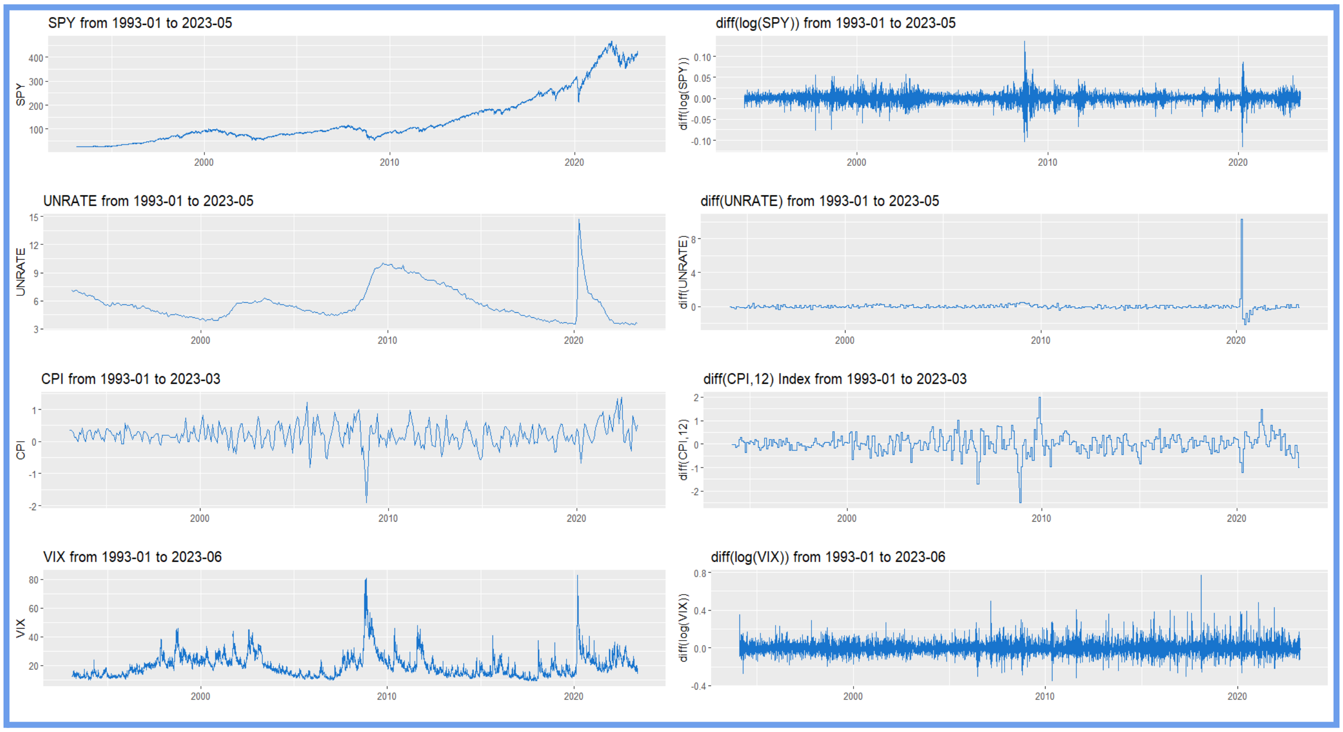
Figure 2: UNRATE shows a significant upward trend between 2008-2009 and in 2020. The series is clearly not stationary, as mean and variance do not stay constant over time.

Figure 3: CPI shows a significant dowward trend around 2008 and an upward trend around 2020. The variance is lower before 2000. The series is not stationary, as mean and variance do not stay constant over time

Figure 4: VIX shows a significant upward trend in 2008 and 2020. Periods of high volatility tend to be clustered together. The series does not have constant mean and variance over time, making it non-stationary.

Transformation to Stationarity

We transform our four time series to stationarity series. Let's compare the before and after of our transformations:



In **Figure 5** (above), SPY series exhibits an upward, increasing trend and unstable variance. Log differencing has removed the trend and reduced the volatility of the series. For UNRATE, because it has a piecewise trend, differencing was done to transform to stationarity. One can notice that there is a spike during the pandemic, which makes sense, since the unemployment rate was affected by the pandemic. Untransformed CPI seems to show some seasonality from the time series plot, so we decided to do seasonal differencing. The series does not have a trend and has more stable variance except around the 2008 recession and the 2020 pandemic. VIX clearly has non-constant variance which we stabilized using log transform. There is somewhat of a piecewise trend too, which we stabilized using a first difference. The choppiness has increased in all plots, which means correlation between lags is not as large as before transformation. We performed the Augmented Dickey–Fuller test for the transformed series and found significant evidence to reject the null hypothesis for each of the series (**Table 1**) which confirms that our series are stationary. *From this point onwards, all our series are used in their stationarized forms.*

Data	*Test statistic*	*p-value*	*Significant evidence to reject the null hypothesis?*
SPY	-19.446	0.01	yes
CPI	-11.039	0.01	yes
UNRATE	-17.699	0.01	yes
VIX	-21.543	0.01	yes

Table 1: Augmented Dickey–Fuller test for transformed SPY, CPI, UNRATE, and VIX.

Model Formulation

Now let us conduct a univariate analysis on the response and predictor variables. We examine our response variable, SPY log daily returns, for its ACF and PACF.

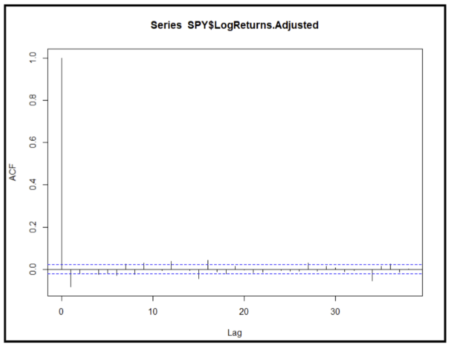


Figure 6: ACF for SPY series appears to be stationary. This appears to be a MA(1) model based on the ACF being zero after lag 1.

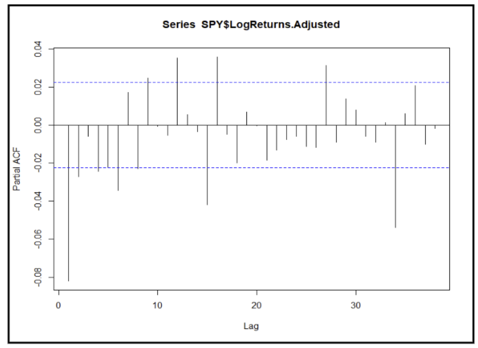


Figure 7: The PACF of SPY log daily returns, given the condition that the series is stationary, cuts off at the early lags rather than tailing off.

We proceed to plot our three predictors' ACF and PACF graphs in **Figure 8** (below).

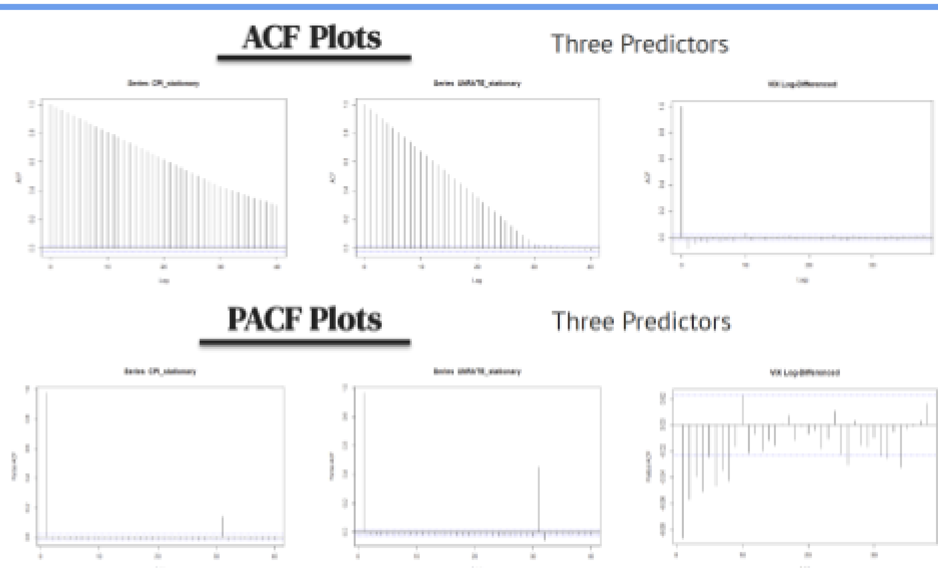
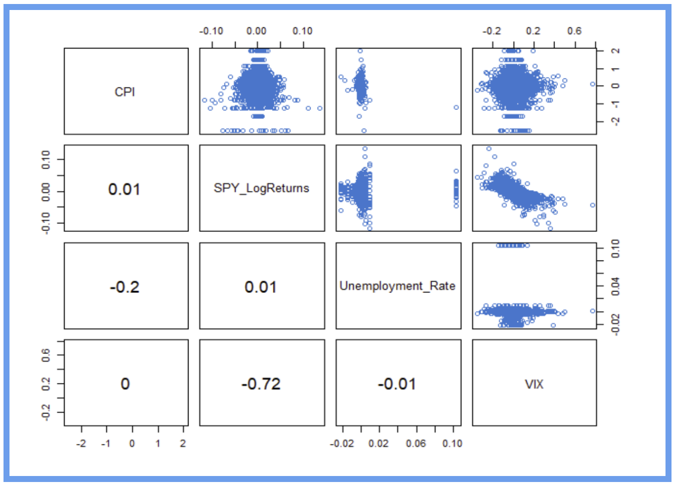
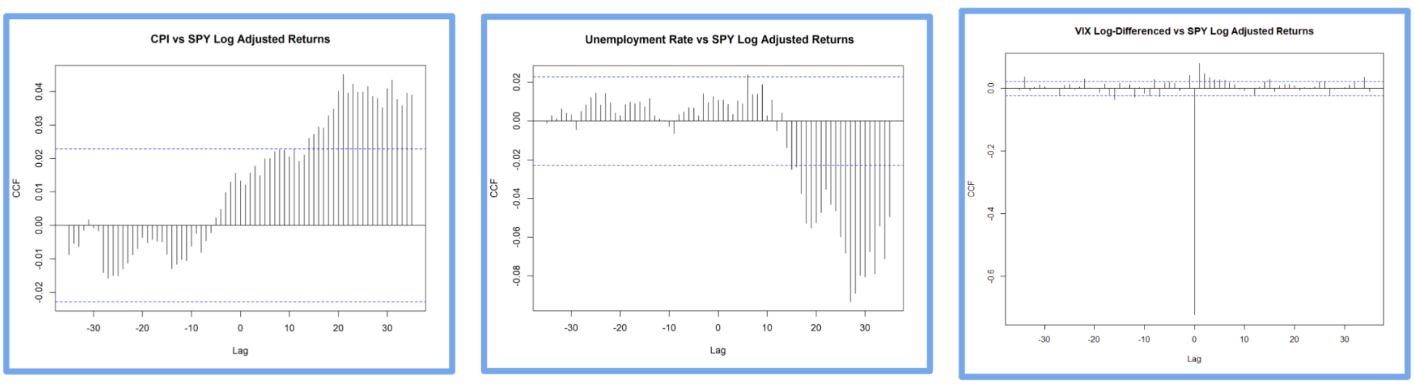


Figure 8: For CPI and UNRATE, the ACFs are tailing off and the PACFs cut off after lag 1. Therefore these are AR(1) processes. For VIX, the ACF cuts off after lag 2, and tails off in PACF. Thus, this is an MA(2) process.

Next we will investigate the relationships between CPI, UNRATE, VIX and SPY to determine if there might be some linear or nonlinear relationships between the features.



In the scatterplot matrix (**Figure 9**, above) there appears to be small magnitude correlations between the predictors, so we will not classify CPI or UNRATE as real-time indicators for changes in SPY, but together with VIX, which has a correlation of -0.72 with SPY, the model may be more viable. UNRATE appears to have a nonlinear relationship with SPY. CPI does not seem to have a significant relationship with SPY. VIX has a strong negative linear relationship with SPY, and it may be the most important predictor in our models.

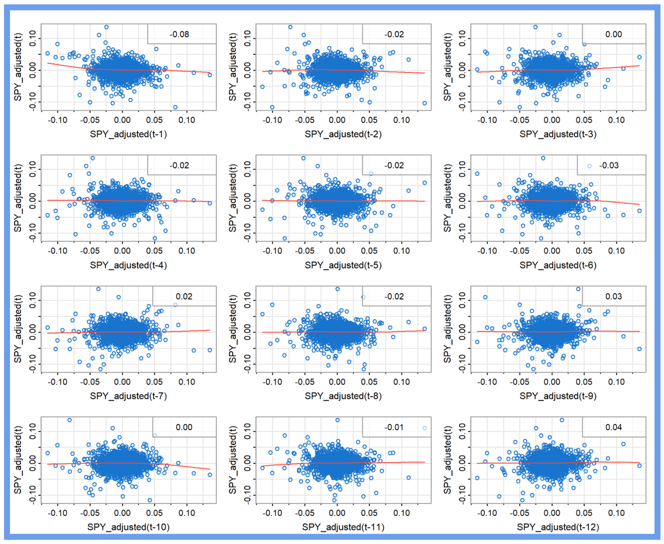


In **Figure 10** above we plot the CCF between pairs of variables to see if we can observe some anomalies, and further, examine whether previous lags will be more interesting.
There are two large spikes in the CCF of CPI vs SPY at lags 21 and 31; this means that the CPI measure at times t+21 days and t+31 days is associated with the SPY series at time t, indicating that the SPY leads the CPI series by 21 and 31 days. The CPI does not lead SPY, which makes sense since CPI is a lagging indicator of inflation. When the stock market is experiencing a downturn, then people are willing to spend less of their disposable income, and try to save money on even essentials; this results in a decrease in CPI.

Also, there is a large spike in the CCF at lag 27 relating SPY to UNRATE. This indicates that the UNRATE measured at time t +27 days is associated with the SPY series at time t, revealing that the SPY leads the UNRATE series by 27 days. The reason why the UNRATE does not lead SPY may be due to the general thought that if many top 500 US companies are not doing well, then this can be indicative of the entire market not doing well, resulting in more layoffs and thus greater unemployment rate. Thus both CPI and UNRATE have large lag relationships with SPY leading them; due to their magnitude it is best that we leave these lags out of the model. Our third predictor variable, VIX, seemed to have stronger cross-correlation with SPY in our scatterplot. We see now in our

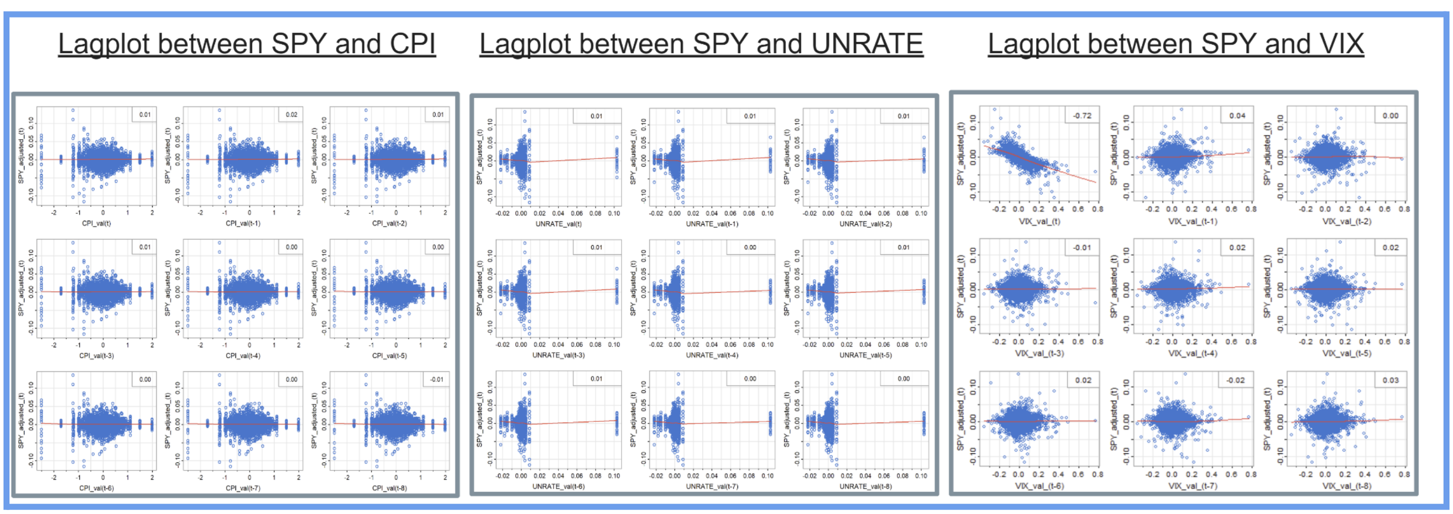
CCF between VIX and SPY that the greatest spike the cross-correlation has is at lag 0, indicating that VIX may be an important predictor for changes in SPY.

Next we will generate lagplots to explore possible linear and nonlinear relations between series at different lags.



In the lagplot between SPY and its previous lags (**Figure 11**, above), we note that all autocorrelations between SPY and previous lags are relatively small. By itself the small autocorrelations seem negligible, but previous SPY lags like lag 1, which has the greatest magnitude, may be useful in future models in combination with other predictors.

In below lagplots (**Figure 12**) for predictors, SPY and CPI have minimal cross-correlation, but CPI at lag 1 offers the highest cross-correlation. Also, SPY and UNRATE are minimally cross-correlated, though possibly nonlinearly related. On the other hand, there is a strong negative cross-correlation between SPY and VIX: -0.72 at lag 0, which makes it seem like VIX is a real-time indicator of SPY.



We began our model-fitting phase by testing a number of regression models with various combinations of our variables to see how our variables predict SPY together (**Tables 2, 3 below**). We also included a t=trend variable to check that we successfully detrended our data. In Models R1-R3 where we tested t, we find that t is highly insignificant, indicating that the differencing was a success.

<u>Regression Tables</u> (Tables 2, 3)

As we performed stepwise variable selection, we noted that the tested models produced neither an acceptable adjusted R^2 value nor significant predictors, at the $\alpha = 0.05$ level. However, the introduction of our VIX variable in Model R7 did produce a moderately good adjusted $R^2 = 0.5288$ with a significant omnibus F-test and two significant predictor variables. R7 also has the lowest AIC and BIC values that we have seen thus far. This aligns with our expectations from our lag plot analysis, since $\gamma_{xv}(0) =$ -0.72. We note that UNRATE is still insignificant in R7 and will investigate the performance of this variable further in our ARIMA models.

							p-values						
Model	SSE	MSE	Adjusted R^2	AIC	BIC	Omnibus	t	c_t	c_{t-1}	c_t^2	u_t	u_{t-2}	v_t
R1	1.0714	1.4591×10^{-4}	-4.022×10^{-5}	-44009	-43974	0.4395	0.917	0.174			0.240		
R2	1.0716	1.4594×10^{-4}	-2.101×10^{-4}	-44007	-43973	0.6921	0.941	0.282		0.713			
R3	1.0714	1.4591×10^{-4}	-1.370×10^{-4}	-44007	-43965	0.5589	0.959	0.200		0.591	0.215		
R4	1.0714	1.4591×10^{-4}	9.454×10^{-5}	-44011	-43983	0.2601		0.175			0.240		
R5	1.0713	1.4592×10^{-4}	1.704×10^{-4}	-44004	-43977	0.1969			0.121		0.226		
R6	1.0713	1.4594×10^{-4}	1.381×10^{-4}	-43997	-43970	0.2216			0.127			0.268	
R7	0.5134	6.9924×10^{-5}	0.521	-49403	-49369	2.2×10^{-16}		0.033			0.625		2×10^{-16}

Model	Model Equation
R1	$x_t = 4.305 \times 10^{-4}_{(6.626\times10^{-4})} - 4.784 \times 10^{-9}_{(4.586\times10^{-8})} t + 4.814 \times 10^{-4}_{(3.539\times10^{-4})} c_t + 2.796 \times 10^{-2}_{(2.378\times10^{-2})} u_t$
R2	$x_t = 4.278 \times 10^{-4}_{(6.628\times10^{-4})} - 3.386 \times 10^{-9}_{(4.607\times10^{-8})} t + 3.774 \times 10^{-4}_{(3.510\times10^{-4})} c_t - 1.135 \times 10^{-4}_{(3.080\times10^{-4})} c_t^2$
R3	$x_t = 4.244 \times 10^{-4}_{(6.628\times10^{-4})} - 2.360 \times 10^{-9}_{(4.608\times10^{-8})} t + 4.570 \times 10^{-4}_{(3.568\times10^{-4})} c_t - 1.674 \times 10^{-4}_{(3.111\times10^{-4})} c_t^2 + 2.977 \times 10^{-2}_{(2.402\times10^{-2})} u_t$
R4	$x_t = 3.630 \times 10^{-4}_{(1.410\times10^{-4})} + 4.804 \times 10^{-4}_{(3.537\times10^{-4})} c_t + 2.797 \times 10^{-2}_{(2.378\times10^{-2})} u_t$
R5	$x_t = 3.633 \times 10^{-4}_{(1.410\times10^{-4})} + 5.480 \times 10^{-4}_{(3.536\times10^{-4})} c_{t-1} + 2.877 \times 10^{-2}_{(2.376\times10^{-2})} u_t$
R6	$x_t = 3.622 \times 10^{-4}_{(1.411\times10^{-4})} + 5.400 \times 10^{-4}_{(3.535\times10^{-4})} c_{t-1} + 2.633 \times 10^{-2}_{(2.375\times10^{-2})} u_{t-2}$
R7	$x_t = 3.714 \times 10^{-4}_{(9.764\times10^{-5})} + 5.226 \times 10^{-4}_{(2.449\times10^{-4})} c_t + 8.049 \times 10^{-3}_{(1.646\times10^{-2})} u_t - 0.128_{(1.430\times10^{-3})} v_t$

<u>ARIMA Tables</u> (Tables 4, 5)

Model	Model Equation
ARIMA(1,0,0).1	$x_t = -0.0830_{(0.0116)} x_{t-1} + w_t$
ARIMA(0,0,1).1	$x_t = -0.0871_{(0.0119)} w_{t-1} + w_t$
ARIMA(1,0,1).1	$x_t = 0.4231_{(0.1641)} x_{t-1} - 0.5038_{(0.1569)} w_{t-1} + w_t$
ARIMA(1,0,0).2	$x_t = -0.0835_{(0.0116)} x_{t-1} + 4.8061 \times 10^{-4}_{(3.2613\times10^{-4})} c_{t-1} + 0.0280_{(0.0219)} u_{t-1} + w_t$
ARIMA(0,0,1).2	$x_t = 4.8076 \times 10^{-4}_{(3.2250\times10^{-4})} c_{t-1} + 0.0280_{(0.0217)} u_{t-1} - 0.0876_{(0.0119)} w_{t-1} + w_t$
ARIMA(1,0,1).2	$x_t = 0.4365_{(0.1586)} x_{t-1} + 4.8985 \times 10^{-4}_{(3.0431\times10^{-4})} c_{t-1} + 0.0286_{(0.0205)} u_{t-1} - 0.5173_{(0.1514)} w_{t-1} + w_t$
ARIMA(1,0,0).3	$x_t = -0.0826_{(0.0116)} x_{t-1} + 5.2279 \times 10^{-4}_{(2.611\times10^{-4})} c_{t-1} + 0.0080_{(0.0152)} u_{t-1} - 0.1277_{(0.0014)} v_{t-1} + w_t$
ARIMA(0,0,1).3	$x_t = 5.2287 \times 10^{-4}_{(2.2361\times10^{-4})} c_{t-1} + 0.0080_{(0.0150)} u_{t-1} - 0.1277_{(0.0014)} v_t t - 1 - 0.0868_{(0.0119)} w_{t-1} + w_t$
ARIMA(1,0,1).3	$x_t = 0.2489_{(0.1351)} x_{t-1} + 5.3569 \times 10^{-4}_{(2.1747\times10^{-4})} c_{t-1} + 0.0074_{(0.0146)} u_{t-1} - 0.1279_{(0.0014)} v_{t-1} - 0.3342_{(0.1318)} w_{t-1} + w_t$
ARIMA(1,0,0).4	$x_t = -0.0819_{(0.0116)} x_{t-1} - 0.1277_{(0.0014)} v_{t-1} + w_t$
ARIMA(0,0,1).4	$x_t = -0.1277_{(0.0014)} v_{t-1} - 0.0859_{(0.0119)} w_{t-1} + w_t$
ARIMA(1,0,1).4	$x_t = 0.2354_{(0.1359)} x_{t-1} - 0.1279_{(0.0014)} v_{t-1} - 0.3198_{(0.1328)} w_{t-1} + w_t$

Following a similar process as with our regression models, we fit a number of AR(1), MA(1), and ARIMA(1,0,1) models based on the analysis of ACFs, PACFs and CCFs explained earlier. While each of our models showed at least one significant predictor, our third ARIMA(1,0,1) model with predictors SPY_{t-1}, w_{t-1}, CPI, UNRATE, and VIX produced the best AIC value, but the BIC value penalizes the higher quantity of variables in this model. Also, our UNRATE variable is, as seen in our R7 model, not significant, along with SPY_{t-1}.

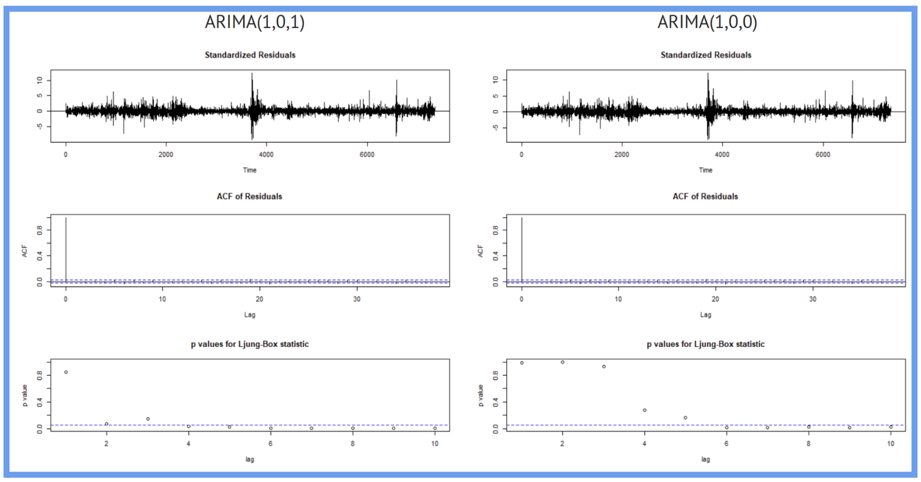
			p-values				
Model	AIC	BIC	SPY_{t-1}	w_{t-1}	CPI	$UNRATE$	VIX
ARIMA(1,0,0).1	-44060.62	-44039.91	9.522×10^{-13}				
ARIMA(0,0,1).1	-44063.07	-44042.37		2.274×10^{-13}			
ARIMA(1,0,1).1	-44065.93	-44038.32	9.927×10^{-3}	1.325×10^{-3}			
ARIMA(1,0,0).2	-44059.87	-44025.37	-7.137×10^{-13}		0.141	0.202	
ARIMA(0,0,1).2	-44062.41	-44027.9		1.635×10^{-13}	0.136	0.197	
ARIMA(1,0,1).2	-44065.65	-44024.25	5.9224×10^{-3}	6.314×10^{-4}	0.107	0.164	
ARIMA(1,0,0).3	-49458.9	-49417.5	1.247×10^{-12}		0.021	0.597	2.2×10^{-16}
ARIMA(0,0,1).3	-49461.49	-49420.08		2.729×10^{-13}	0.019	0.593	2.2×10^{-16}
ARIMA(1,0,1).3	-49463.08	-49414.77	0.065	0.011	0.014	0.615	2.2×10^{-16}
ARIMA(1,0,0).4	-49457.49	-49429.88	1.930×10^{-12}				2.2×10^{-16}
ARIMA(0,0,1).4	-49459.94	-49432.34		4.527×10^{-13}			2.2×10^{-16}
ARIMA(1,0,1).4	-49461.17	-49426.67	0.084	0.016			2.2×10^{-16}

We examined our models and determined that our fourth AR(1) model is currently the most optimal model. Not only are both predictors, SPY_{t-1} and VIX, highly significant, this model is much simpler than the ARIMA(1,0,1) model, while having comparable AIC and BIC diagnostic values and interpretable variables. We thus determine that an AR(1) model using our SPY_{t-1} and VIX is our best choice at present.

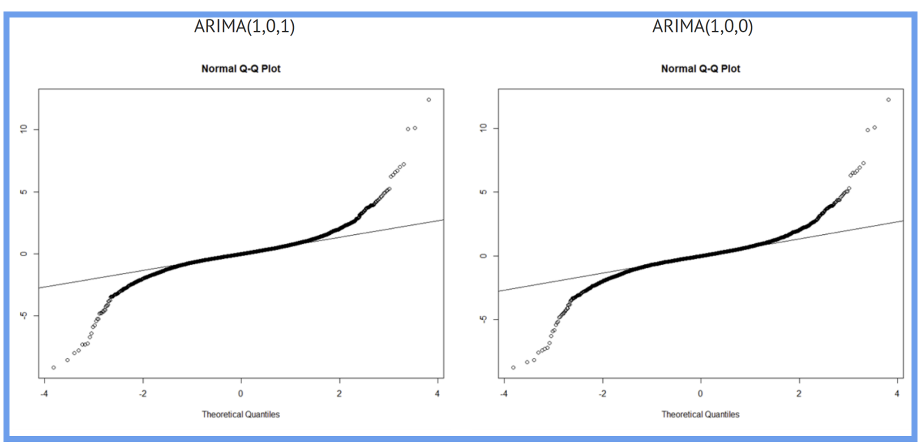
We computed Yule-Walker estimates for our models, and since we have such a large sample size, the results were nearly identical to those of our ARIMA models. We also ran an Auto-ARIMA that closely matched our chosen model. To solidify our decision, we examine the residual diagnostic plots for the two optimal ARIMA models.

Diagnostics

Diagnostic plots (**Figure 13**, below) show that ACFs closely resemble white noise for both models. The Ljung-Box test for ARIMA(1,0,1) rejects the null hypothesis which means the residuals are highly correlated at lags 2 and beyond, and it rejects it for ARIMA(1,0,0) for lags 6 and beyond. The residuals exhibit a high degree of variance for both models at multiple time points. We conclude that the residuals for ARIMA(1, 0, 0) more closely resemble white noise, making it a slightly better model.



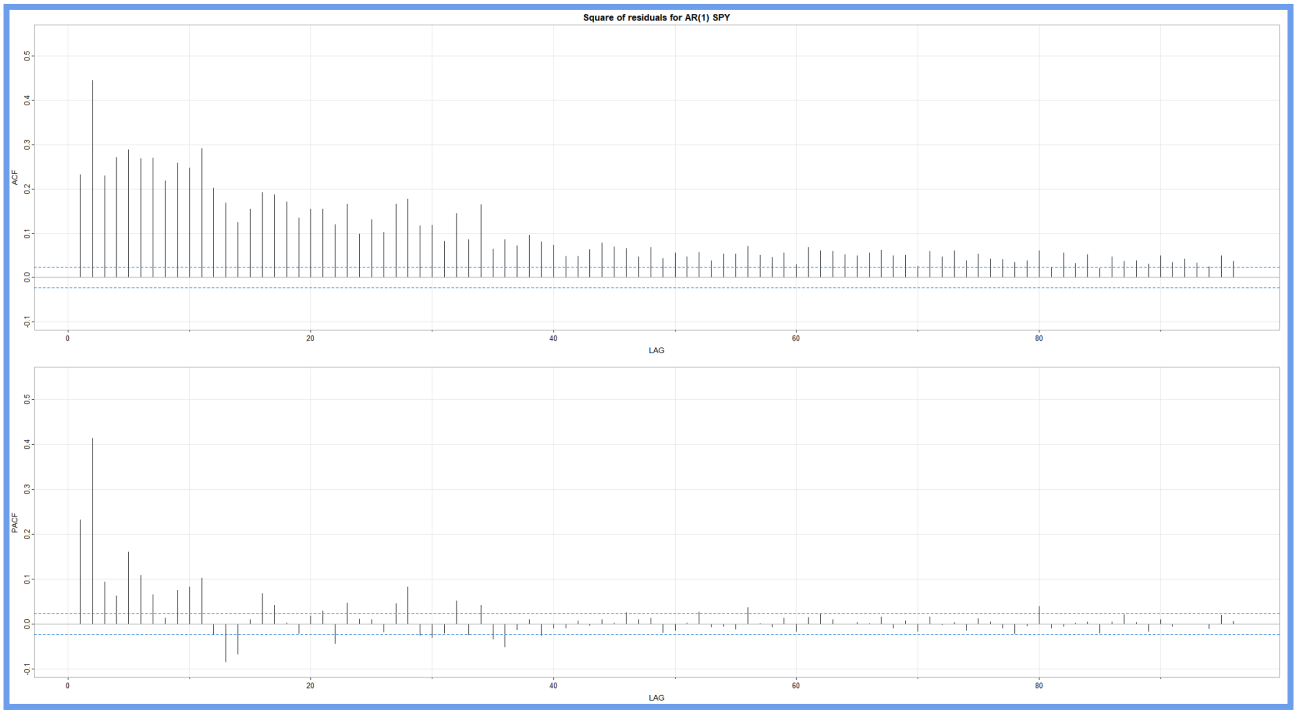
Q-Q plots (**Figure 14**, below) of the two models clearly show that the residuals are not normal. This is due to the fatter tails, likely because of the global events during 2008 and 2020 leading to outliers.



Further investigation needs to be done to get the distribution of the model residuals closer to normal, e.g., we can explore adding an indicator variable for pandemic status and check the significance of the interaction term in the model, or consider an ARCH model for the variance issue. Overall, ARIMA (1, 0, 0) looks like the better model because it's simpler and more interpretable: SPY is intuitively forecasted by its previous lag and the current value of the VIX index.

Note: since we did not observe any seasonality in our data, we do not fit SARIMA models to the series.

A/GARCH Models



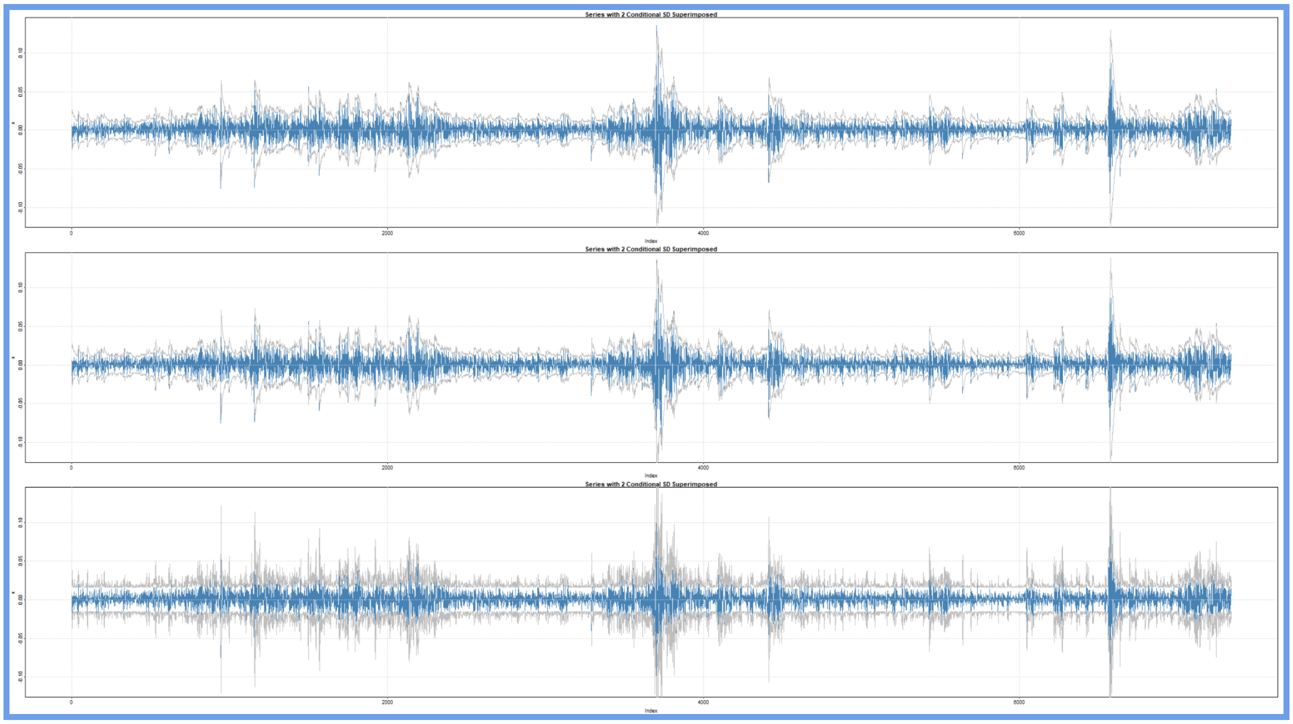
P/ACF of the squared residuals from the AR(1) SPY model show a high degree of correlation indicating that a G/ARCH model must be fit. Lag 2 has the largest correlation in both plots. Both plots decay after lag 2 but there are significant correlations at higher lags as well.

Model	AIC	BIC
AR(1) - GARCH(1,1)	-6.497366	-6.491726
AR(1) - GARCH(2,2)	-6.500073	-6.492554
AR(1) - GARCH(2,0)	-6.387835	-6.382196

Table 5: AR-GARCH Models

Model	AIC	BIC
AR(1) - GARCH(1,1)	-6.497366	-6.491726
AR(1) - GARCH(2,2)	-6.500073	-6.492554
AR(1) - GARCH(2,0)	-6.387835	-6.382196

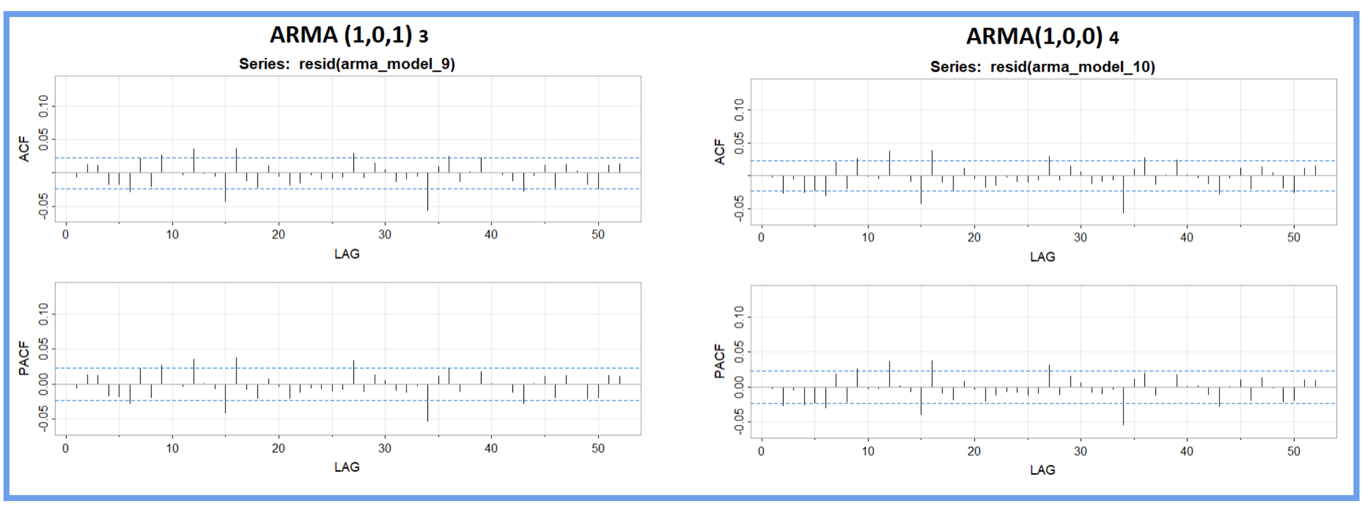
We decided to try multiple GARCH models with the AR(1) SPY model: AR(1)-GARCH (2,2) is chosen because both ACF and PACF have a significant correlation at lag 2. AR(1)-GARCH (2,0) is chosen because the PACF seems to cut off after lag 2 while the ACF decays. AR(1)-GARCH (1,1) is chosen as the simplest and most parsimonious model since lag 1 correlation is significant in P/ACF.



To explore the AR-GARCH predictions of volatility, we calculated and plotted the SPY log returns along with one-step ahead predictions of the corresponding volatility. The top plot is for AR(1)-GARCH (1,1), middle plot is for AR(1)-GARCH (2,2) and the last plot is for AR(1)-GARCH (2,0) (which is really an AR(1)-ARCH(2) model). The more complicated models, AR(1)-GARCH(1,1) and AR(1)-GARCH (2,0), capture the finer changes in volatility. AR(1)- GARCH (2,2) has the lowest AIC/BIC score which makes it the best AR-GARCH model.

Regression with Auto-Correlated Errors

In the above fitted ARIMA models, it follows the classical regression assumption of error terms being uncorrelated. If the error terms are not white noise, and there seems to be a pattern, we can follow the approach of fitting an AR/MA/ARMA model to the residuals and transforming the regression equation to a model that has its error terms uncorrelated. Therefore, we take the best two ARIMA models and try to check for correlation about the error terms.



Evidently, the residuals analysis shows no obvious departure of the residuals from white noise.

Model	*RMSE*
ARIMA (1,0,1)	0.008330008
ARIMA (1,0,0)	0.008336583
AR (1)-GARCH (2,2)	0.01207451

Table 6: Root Mean Squared Error (RMSE) for ARIMA (1,0,1), ARIMA (1,0,0) and AR (1)-GARCH (2,2).

Forecasting and Prediction

We performed a 5-days ahead forecast using the ARIMA(1,0,1), ARIMA(1,0,0) and AR(1)-GARCH(2,2) models and compared their Root Mean Squared Errors (**Table 6**). ARIMA (1,0,1) has the lowest RMSE, so we choose this over the AR(1) model for prediction purposes. The point forecast and 95% prediction intervals for ARIMA(1,0,1) model are given in **Table 7**.

Point Estimate	Lower Bound	Upper Bound
0.0007540953	-0.01557272	0.01708091
0.0016662239	-0.01471902	0.01805146
0.0019077961	-0.01448153	0.01829713
0.0019717752	-0.01441784	0.01836139
0.0019887197	-0.01440092	0.01837836

Table 7: Prediction Intervals for SPY Log Daily Returns using the ARIMA (1,0,1) model

Back Transformation

We back transformed our forecasts to SPY values for easier interpretation.

Let x_n denote the transformed series at time n. Let spy_n denote the original SPY series at time n. The m-step ahead forecast for the transformed series is given by:

$x^n_{n+m} = E[x_{n+m} | x_n ... x_1]$.

We can now express x^n_{n+m} as:

$x^n_{n+m} = \mathrm{E}[log(spy)_{n+m} - log(spy)_n \mid x_n ... x_1]$ … (1).

We know the value of $log(spy)_n$ because we have the knowledge base till time n. Using this fact, (1) can be written as:

$x^n_{n+m} = \mathrm{E}[log(spy)_{n+m} \mid x_n ... x_1] - log(spy)_n = log(spy)^n_{n+m} - log(spy)_n \Rightarrow log(spy)^n_{n+m} = x^n_{n+m} + log(spy)_n$

$\Rightarrow (spy)^n_{n+m} = exp(x^n_{n+m} + log(spy)_n) = spy_n * exp(x^n_{n+m})$ .. (2).

We used (2) to back transform the forecasts (**Table 8**) for ARIMA (1,0,1) and compared them with the original SPY values. The forecasted values are not far off from the original values indicating that our model has reasonable accuracy.

Date	Actual	Prediction Point Estimate	Prediction Lower Bound	Prediction Upper Bound
April 3rd , 2023	409.421	408.1826	401.5724	414.9016
April 4th , 2023	407.1576	408.8633	395.7049	422.4592
April 5th , 2023	406.0916	409.6441	390.0158	430.2602
April 6th , 2023	407.6756	410.4526	384.4330	438.2333
April 10th , 2023	408.0941	411.2697	378.9364	446.3618

Table 8: Actual, Forecast SPY values, and Prediction Intervals from April 3rd, 2023 – April 10th, 2023 (Excluding public holidays and weekends)

Conclusion

The goal of this paper was to predict future SPY daily returns over 1993-01-29 to 2023-03-31, a 30 year period that has seen its fair share of ups and downs in the stock market. There have been notable world events that have inspired “bursts of volatility” in the time series, such as the dot com bubble burst in 1999-2000, the 2008 housing market crisis, and 2020’s pandemic market downshift, which have indubitably left their mark on the financial landscape. Despite these troubles, the shine of the stock market has not dulled, with the motto of “greater risk yields greater reward” still echoing on Twitter and chat groups. In our analysis, we proceeded step by step, from Regression to ARIMA to GARCH, to uncover the models that can help the average investor make more informed investment decisions in the SPDR ETF (SPY), and thus the greater S&P 500 index.

We discovered in our study that the ARIMA models produced a prediction set that was altogether more appealing than that of the GARCH models because we are modeling the daily returns; however, we would expect GARCH to be more appealing if we were modeling the variance of daily returns. AR(1) is overall a better model in terms of interpretability, but ARIMA(1,0,1) was found to be better for forecasting due to its comparatively lower RMSE value.

The impact of modeling the volatility of the S&P 500 and predicting future returns can be vast. By being able to predict the future return of stocks tomorrow, or next week, or perhaps even next month, traders can make more informed decisions about buying, holding and selling. This can in turn lead to periods of economic growth, since if the average person can more optimally predict what stocks will perform better, then the market can experience even more buying and selling activities. In addition, the ability to accurately model the volatility of the S&P 500 through VIX and GARCH models can lead to improved precision in option pricing, since traders can target specific stocks for maximum profit in a timely manner. In the future, we can improve our prediction of volatility by using GARCH models that may be more specific to our data like APARCH, EGARCH, and other methods.

By allowing the average person to more affordably invest in the stock market, the ETF SPDR breaks down gatekeeping in the S&P 500. Therefore, with the addition of models that lead to greater yield and growth, the public gains the tools to take their financial destiny into their own hands.

References:

1. Yahoo! (n.d.). What is the adjusted close? | yahoo help - SLN28256. Yahoo! https://help.yahoo.com/kb/SLN28256.html#:~:text=Adjusted%20close%20is%20the%20closing,Security%20Prices%20(CRSP)%20standards.

2. Investopedia. (n.d.). Financial terms dictionary. Investopedia. https://www.investopedia.com/financial-term-dictionary-4769738

3. Doran, James and Ronn, Ehud I. and Goldberg, Robert S., A Simple Model for Time-Varying Expected Returns on the S&P 500 Index (October 31, 2008). Journal of Investment Management, Forthcoming, Available at SSRN:https://ssrn.com/abstract=788825

4. Chicago Board Options Exchange. (n.d.). Cboe VIX Index. VIX Index. https://www.cboe.com/tradable_products/vix/

5. BetterTrader.co. (n.d.). *Case study - unemployment rate and its effect on the S&P 500: It's less intuitive than you think*. BetterTrader online trading tool. https://bettertrader.co/case-studies/unemployment-rate-and-its-effect-on-the-SP-500.html

6. Ionides, E. (2018, March 4). *Time series analysis for log returns of S&P500 - GitHub Pages*. Time Series Analysis for Log Returns of S&P500. https://ionides.github.io/531w18/midterm_project/project38/Midterm_proj.html

7. Kumar, M. (2023, April 26). *Correlating Vix and spy: A statistical study using Python's linear regression*. Medium. https://medium.com/@maneesh1.kumar/correlating-vix-and-spy-a-statistical-study-using-pythons-linear-regression-9261bfd9dd71

8. Rob J Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice. 2nd edition*. Monash University, Australia

9. Winberg, A., & Rugås, N. (2015, February 24). *A study on the relation between VIX, S&P500 and the CDX-index* . NEG-300 H14 Project Paper with Discussant – Finance (15 ECTS). https://core.ac.uk/download/43558302.pdf