ARIMA, GARCH, K-NN on Stock Price Prediction

Benjamin Ryu*

December 10, 2024

Abstract

This paper looks into the predictability of stock prices using ARIMA, GARCH, and K-NN models, focusing on their ability to forecast stock movements. Drawing on methodologies from prior research, we applied these models to a dataset comprising four stock tickers—SPY, AAPL, NVDA, and GLD erived from Yahoo Finance, spanning October 1, 2023, to October 1, 2024. Preliminary results indicate varied performance, with the K-NN model demonstrating notable accuracy in predicting SPY's trends, while the ARIMA model faced limitations in capturing volatility, while the GARCH model was able to capture shock. This work contributes to the broader understanding of machine learning and econometric applications in financial forecasting, highlighting the importance of model selection and parameter tuning for improved accuracy.

Keywords: Stocks; Forecasting

JEL Codes: C53, C58, G13

*Department of Economics, San José State University

1 Introduction

The stock market tends to be hard to predict, many attempts are made by retail traders often not being able to beat the overall market performance. This article aims to use different forecasting and machine learning methods to find the optimal method of stock market prediction. We understand that there are economic indicators that affect the direction of the average market. However individual stocks may not be as sensitive to the economic indicator, therefore variables may not weigh as heavily on the direction of the stock market. This is why a lot of forecasts use Random Walk which we will also test.

Our models are derived from several past literature surrounding the same subject. We will be deriving the same models to test the predictability of the models. Ogundunmade TP (2022) and Malagrino et al. (2018) are advanced papers that do not necessarily use time series to make their prediction but more advanced statistical tools. While Saha and Bose (2020) uses a more classical model along with an advanced model to predict the market. Our models will look after these past literature and will be heavily influenced by their works and results.

Our data is derived from Yahoo Financial, we will be looking at index funds, Apple, Nvidia, and gold funds. The reason behind looking at multiple stocks is to test models under different performing stocks. While these stocks are all long-term stocks that are heavily tied to economic conditions we are expecting to see the model react differently among them as the predictability and the implied volatility remain different.

In this article, we are going to use 3 different models. The first one is Autoregressive integrated moving average or ARIMA, this model is going to be the most basic model we have. It will work as the bench test for the rest of our models. Next, we have Generalized AutoRegressive Conditional Heteroskedasticity or GARCH, this model is similar to the ARIMA model but the variance of the error term is serially autocorrelated following an autoregressive moving average process. Our next model is K Nearest Neighgbor or K-NN which is our machine learning model and will be using a non-parametric supervised learning.

We found that the K-NN model works best according to rolling origins but ARIMA result closely tracked the result of actual values.

2 Literature Review

Ogundunmade TP (2022) article attempts to predict Nigerian stock marketing using machine learning models with K-fold and repeated K-fold CVs. Ogundunmade TP (2022) use K-fold and repeated k-fold CVs as measurements of the performance of machine learning models. The authors used a simple linear regression model, random forest (RF), classification and regression tree (CART), an artificial neural network, and the support Vector Machine model. The article uses real gross domestic product, inflation rate, exchange rate, and interest rate in their regression input. They in conclusion found that CVs produce better results than models with no CV technique and that the RF model with CV was the best in their prediction for stock exchange prices in Nigeria.

Malagrino et al. (2018) article much like Ogundunmade TP (2022) looks to predict the stock market, but in particular they look to predict the general direction of the stock market. The authors use a Bayesian Network to forecast the direction of the São Paulo Exchange main index. Malagrino et al. (2018) use indices from the global stock market to predict the next day's closing direction of the São Paulo Exchange main index. This article also introduces CVs to reduce any influence seasonality might have. In conclusion, they found that the model was 71% accurate in predicting the direction of the market which was similar to other literature regarding the same topic.

Saha and Bose (2020) article compares classical time series model with machine learning and single layer neural networks models. Their article uses ARIMA, GARCH, Prophet, K-NN regression time series forecasting, and a feed-forward neural network to predict the stock market. Their article attempt to forecast a horizon of 30 days using the 5 models stated. They found that all the models are overestimating the prices of the stocks. Saha and Bose

(2020) found that ARIMA and the neural network models worked best within the parameter compared to the rest of the models. They suspect that the other models are relatively novel at the time of writing and need of more tuning the parameters.

Chatterjee et al. (2021) article looks to predict stock price using ARIMA, Holt-Winters Exponential Smoothing, Random Forest, MARS, RNN, and LSTM. They look into the data of Infosys, ICICI, and SUN PHARMA from the period of January 2004 to December 2019. They split the data into a test set and a training set to see what model works best. The article found that the MARS was the most effective machine learning model and the LSTM to be the best deep learning model.

3 Data

In this section, we will discuss the sources of our data. Our data is derived from Yahoo Financial which keeps track of stock prices from open, close, that day's high, that day's low, volume sold, and lastly adjusted close after all splits and dividend distributions [1]. Specifically, we are looking at a year's worth of data from October 1st, 2023 to October 1st, 2024 which accounts for 251 observations for four stocks(SPY, AAPL, NVDA, GLD). The reason we do not have 365 observations is due to the exclusion of weekends and holidays in which the market does not operate. Our model only uses closing price which is indicated as Close in Tables 1 to 4.

Table 1: Summary Statistic: SPDR S&P 500 ETF Trust

Statistic	N	Mean	St. Dev.	Min	Max
Open	251	504.988	42.961	413.560	574.380
High	251	507.292	42.975	414.600	574.710
Low	251	502.534	42.615	409.210	570.420
Close	251	505.155	42.775	410.680	573.760
Volume	251	66,425,693	22,480,823	27,289,700	146,267,400
Adjusted	251	501.508	44.254	405.201	573.760

[1] https://finance.yahoo.com/

The first summary statistic(Table 1) is SPDR S&P 500 ETF Trust more commonly known as SP500 or SPY. The SPY is an index fund that holds a portfolio comprising all 500 companies on the index. This company was picked by State Street Global Advisors to reflect the market-performing companies. So often the SPY will have companies in many different industries and will be affected differently by economic conditions. This sets a good benchmark stock as the implied volatility should be lower the most individual stocks.

Table 2: Summary Statistic: Apple Inc.

Statistic	N	Mean	St. Dev.	Min	Max
Open	251	194.074	20.023	165.350	236.480
High	251	195.881	20.257	166.400	237.230
Low	251	192.445	19.590	164.080	233.090
Close	251	194.283	19.946	165.000	234.820
Volume	251	59,485,482	29,733,849	24,048,300	318,679,900
Adjusted	251	193.621	20.113	164.405	234.291

Next, we have Apple Inc. as in seen Table 2, is a technology company well-known for its iPhones. For a large majority of the 21st century, they were the most valued company in the United States stock market in terms of total value. Apple is considered a blue-chip stock, which is any stock that is considered a safe long-term investment relative to the stock market in general. Much SPY Apple than to be very stable Apple is part of the SPY 500 companies.

Table 3: Summary Statistic: Nvidia Corporation

Statistic	N	Mean	St. Dev.	Min	Max
Open	251	84.740	30.250	40.450	139.800
High	251	86.356	30.933	40.879	140.760
Low	251	82.934	29.332	39.230	132.420
Close	251	84.742	30.121	40.326	135.580
Volume	251	428,600,273	141,944,330	173,911,000	1,142,269,000
Adjusted	251	84.730	30.122	40.314	135.568

Nvidia Corporation(Table 3) is the highest-value company in the United States at the time of writing this article. They are best known for their graphic processor units, often needed to run computers, and were thrust in market value through the AI revolution in the last couple of years. NVIDIA has recently joined Apple in the SPY 500 list, the company had tremendous growth over the years, and it is considered less stable than Apple due to its sudden growth. This will test the model on the ability to adjust for shocks in market growth.

Table 4: Summary Statistic: SPDR Gold Trust

Statistic	N	Mean	St. Dev.	Min	Max
Open	251	205.243	19.446	168.740	246.440
High	251	206.179	19.578	169.000	247.370
Low	251	204.330	19.210	168.300	245.190
Close	251	205.322	19.429	168.830	246.980
Volume	251	7,357,544	3,243,189	2,445,100	30,876,600
Adjusted	251	205.322	19.429	168.830	246.980

Lastly, the SPDR Gold Trust(Table 4) is a gold trust that closely tracks real gold value. This often acts in the inverse of the dollar and general stock market. It is often used to hedge the market as it does well when the general market is down along with the economy. It is believed to be stable and implied volatile is generally believed to be low.

4 Empirical Methods

We have three empirical methods in this article, Equation(1) is the ARIMA model, Equation(2) and Equation(3) are GARCH model, and lastly Equation(4), Equation(5), and Equation(6) are K-NN model equations. These three models will be used to predict the stock market.

$$y_t = c + \phi_1 y_d t - 1 + \phi_2 y_d t - 2 + \cdots + \phi_p y_{dt-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \tag{1}$$

Equation(1) is mentioned in the ARIMA model where the Auto Regression(AR) depends on past values, Intergrated(I) is differencing the data to make it stationary, and Moving Average(MA) relies on past data. We see that ϕ is the AR coefficients lagged to p while MA coefficients ϵ lagged to q, while d represents the differencing that makes it stationary.

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{2}$$

$$y_t = \mu + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2) \tag{3}$$

$$\hat{y} = \text{mode}\{y_i \,|\, x_i \in \mathcal{N}_k(x)\} \quad \text{(for classification)} \tag{4}$$

Equations(2) and (3) are for GARCH models, the model forecast time series with volatility clustering, which is common in financial data like stock market returns such as our own. Equation(2) shows us the conditional variance at a given period t. The first summation $\sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2$ is the ARCH(Autoregressive conditional heteroskedasticity) where we are looking at the immediate impact of shocks on volatility. While second summation $\sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$ is the GARCH(Generalized autoregressive conditional heteroskedasticity) term part of our model that captures the persistence of volatility, which sets our current volatility is dependent on the previous levels of volatility. Equation(3) is the mean equation where μ is the mean of the time series.

$$\hat{y} = \frac{1}{k} \sum_{x_i \in \mathcal{N}_k(x)} y_i \quad \text{(for regression)} \tag{5}$$

$$d(x, x_i) = \sqrt{\sum_{j=1}^{n} (x_j - x_{i,j})^2} \quad \text{(Euclidean distance)} \tag{6}$$

Equations(5) and (6) are the K-NN models, this model is a machine learning model that uses non-parametric supervised learning. the K-NN model is often used for classification but is also applied in regression. In our article, we are aiming to use this model to find the nearest neighbor of the next value based on the past value. The K-NN model leverages historical data to forecast future values by identifying the nearest neighbors of a given data point based on past values. In this study, we apply the K-NN model to forecast stock prices, using past values to find the nearest neighbors of the next predicted value.

5 Results

In this section we will go over results, particularly we will be going over our ARIMA, GARCH, and K-NN models results. We will be predicting 30 observations in our models. This section is divided into subsections by type of model.

5.1 ARIMA Results

In this subsection, we will go over the results of our ARIMA models. The first model is SPY ARIMA as seen in Table 5 model (1) with the given parameters $(1, 1, 0)$. Our value estimate shows a weak positive relationship between the current and the lag value. In addition, we see that the standard error is large suggesting that the coefficient is not as precise. Our model (1) shows no statistical significance variables with our chosen parameter. We chose our parameter base a function in R called auto.arima that picks optimal parameters. Models (2) and (3) also show similar results as our first model with the same parameters $(1, 1, 0)$. Only our NVDA model (3) used parameter $(2, 1, 2)$. It should be noted that all our models are nonstationary therefore our $d = 1$ in all parameters.
Figure 1 is our SPY ARIMA model which shows a huge variance. Our Table 5 shows that model (1) is not statically significant with the autoregression 1 being 0.041. Our parameter was determined by our finding in Figure 21. We saw that there was a significant spike in all lags in our ACF and a spike in the first lag in our partial ACF.

Our next figure is the ARIMA result for AAPL, as mentioned, the parameter used is identical to our SPY model as seen in Figure 22 and the result is similar to Figure 21. We

Figure 1: SPY

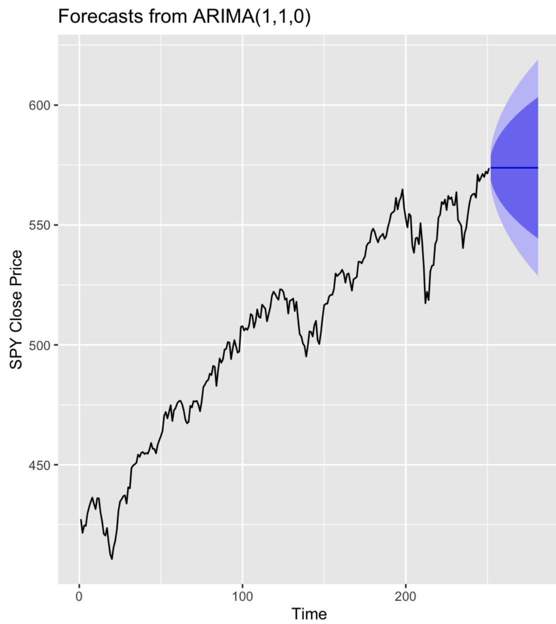
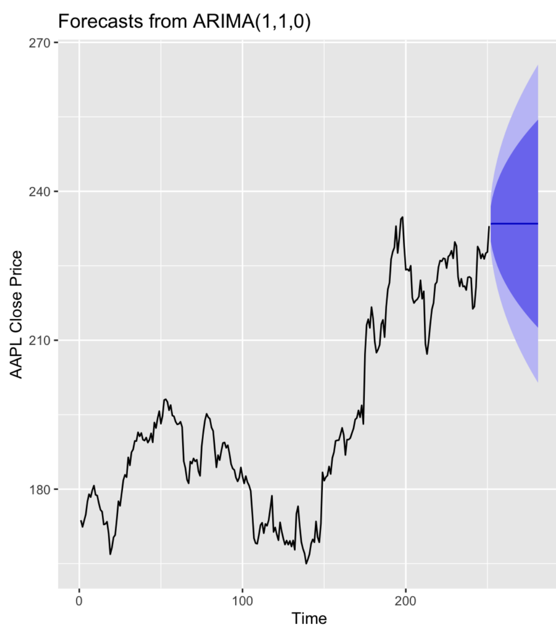


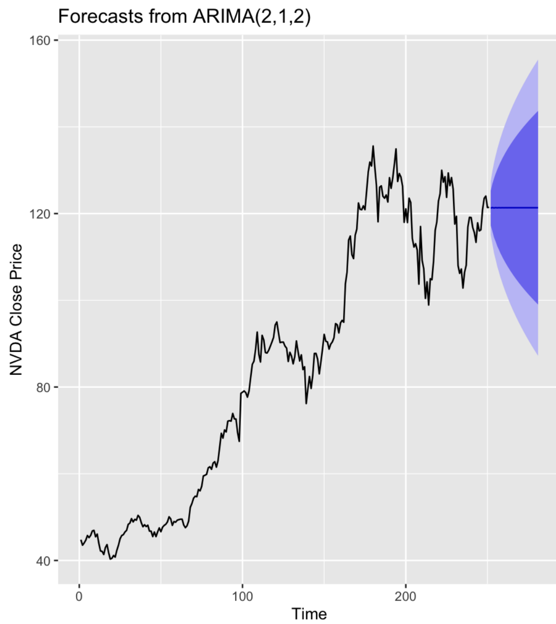
Figure 2: AAPL



see graphically the model has a wider variance than SPY and this is seen with our autoregression result in Table 5 Model (2). We can see that both models show the same standard error but the model for AAPL is even less significant than SPY.

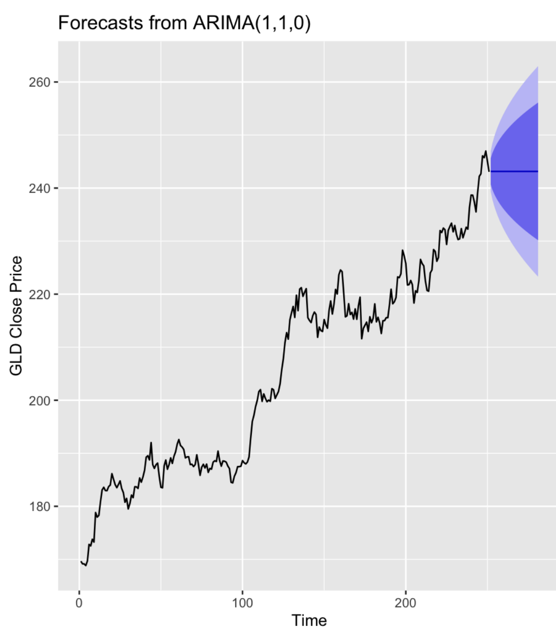
NVDA was our other test for the ARIMA model which deviates from Model (1) and (2)

Figure 3: NVDA



with the parameter $(2, 1, 2)$. This model has 2 lags on the autoregression and moving average which was determined by the partial ACF having two spikes in Figure 23. This means our results have more variables to look at, but all results returned non statically significant.

Figure 4: GLD

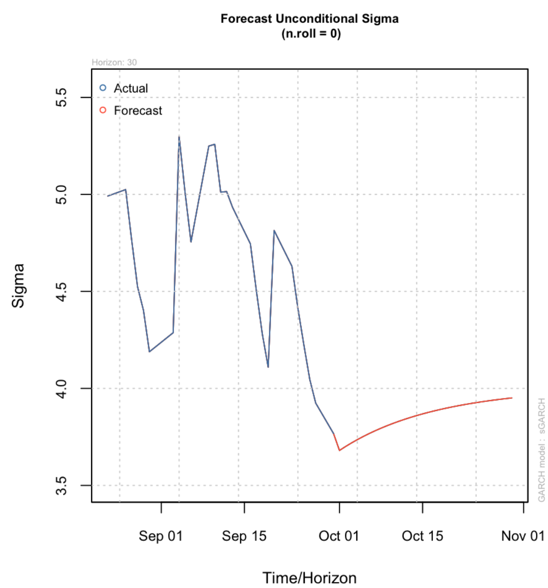


Lastly, we have the GLD has the same parameter as the first 2 models and shows similar results to the SPY. Both stocks are stable stocks they both share similar attributes, unlike AAPL and NVDA. The model was also shown to have a similar statistical significance as the SPY model as shown in Table 5 Model (4). We are not able to come to a conclusion with these results alone.

5.2 GARCH Results

In this subsection, we have our GARCH result for the same 4 stock tickers. Our GARCH results are represented in both forecasting graphs and result plots as seen through Figures 5 to 8 and Figures 25 to 28 respectively.

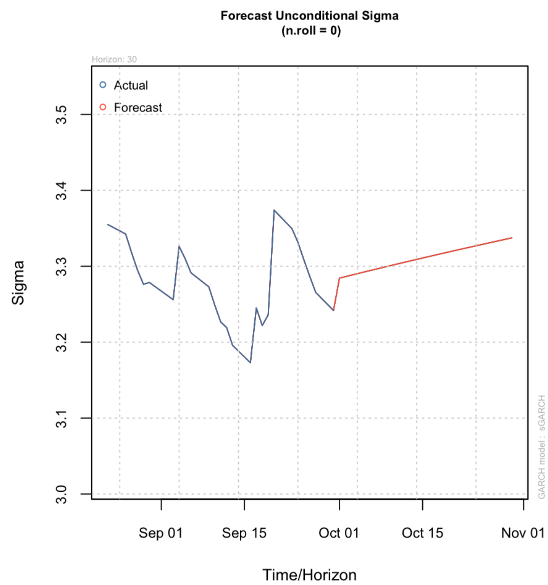
Figure 5: SPY GARCH



Our Figure 5 is the GARCH forecast for SPY with 30 observation forecasting using the Sigma Prediction (unconditional) plot. We see that the forecasting is nonlinear in our results. Our Garch result as seen in Figure 25 sees an absence of residual serial correlation and ARCH effects, supporting the model's adequacy.

Figure 6 shows the same model with AAPL. Again the model shows us a nonlinear forecast using the GARCH Sigma Predictions plot. Based on our GARCH result in Figure

Figure 6: AAPL GARCH



26 we can conclude that The GARCH(1,1) model provides a good fit for AAPL returns, capturing the volatility clustering and the autoregressive nature of the data. Again our diagnostic tests confirm the absence of serial correlation, ARCH effects, or bias, supporting the model's validity much like the SPY GARCH.

Figure 7: NVDA GARCH

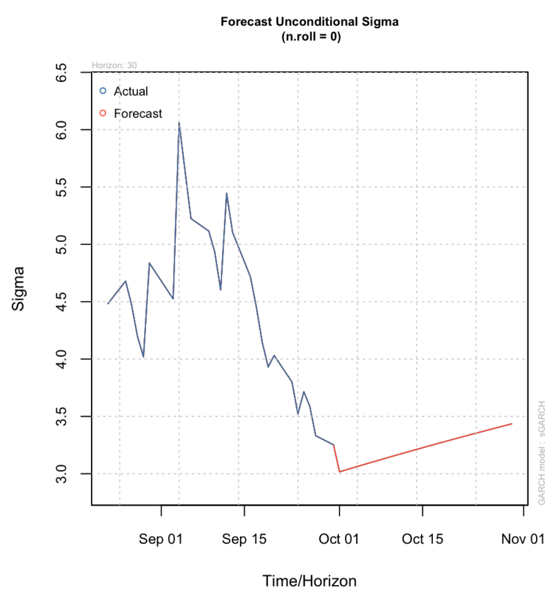
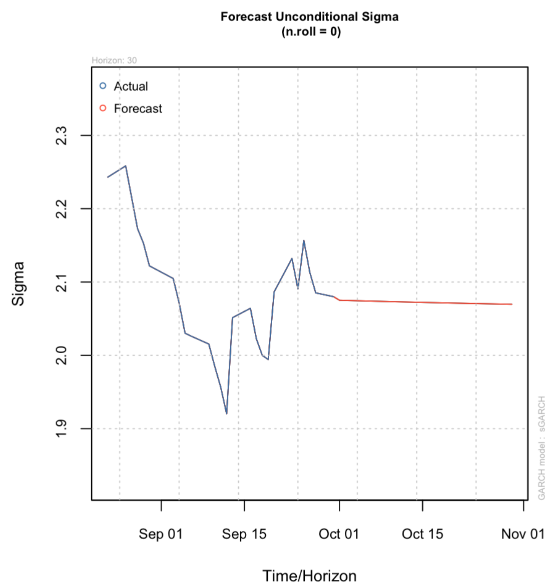


Figure 7 for NVDA shows a forecasting result similar to that of the rest of the other figures. But NVDA unlike the other stock tickers has seen an increase in stock price due to shocks in the market for chip design and manufacturing industry. We can see that in Figure 27, we see that there is an absence of serial correlation, ARCH effects, and sign bias, supporting the model's adequacy but the the ar1 has an instability which suggests the potential variability in the returns' autoregressive structure.

Figure 8: GLD GARCH



Lastly, we look at Figure 8 which is for GLD. There is nothing of note to mention in the figure as the forecast looks standard. Our diagnostic result in Figure 28 shows supports the model's adequacy, with no major issues in residual correlation, ARCH effects, or parameter stability. In conclusion, our model added on the ARIMA by adjusting for any possible shock as seen with our NVDA model, but the result is still not robust enough to conclude the effectiveness of the model.

5.3 K-NN Results

In this subsection, we will discuss the result of our K-NN model. The K-NN subsection will look over the forecasting result of our 4 tickers. K-NN model like other models forecasts 30 obersartion. One of the other inputs for this model is the k value which we are going to use Saha and Bose (2020) k value 50.

Figure 9: SPY K-NN

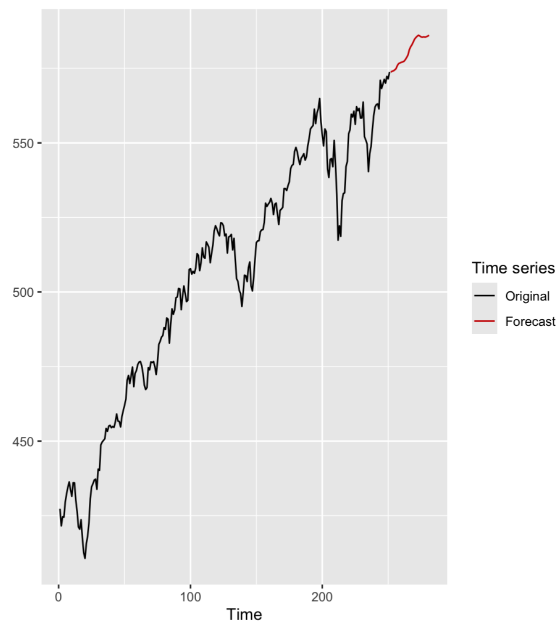
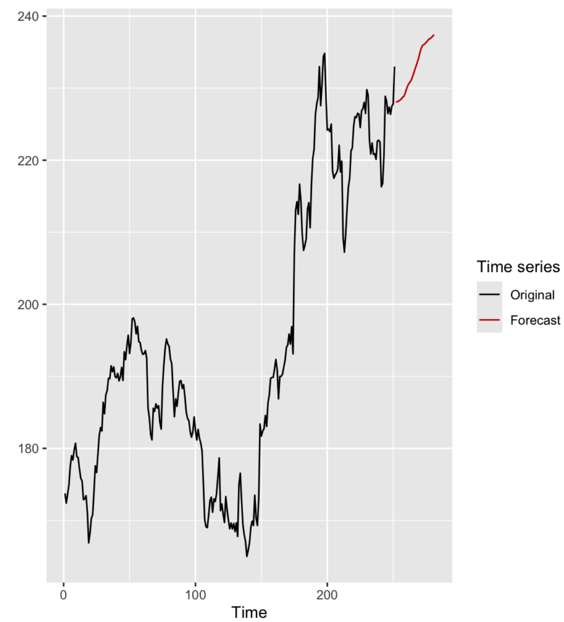


Figure 9 is our SPY K-NN forecast, we see that the forecast seems to be more accurate than our other forecast. Visuals seem to follow the overall trend line. However, it is hard to determine the accuracy of the model without data to compare with. We will further test this in our robustness check of this model. We however will discuss the rolling origins result of the K-NN model as seen in Table 6. The result indicated that the low values of RMSE, MAE, and MAPE collectively suggest that the K-NN model provides accurate forecasts for the SPY time series with minimal error. The MAPE being just above 1% highlights the model's reliability, especially in terms of proportional error relative to the magnitude of the data.

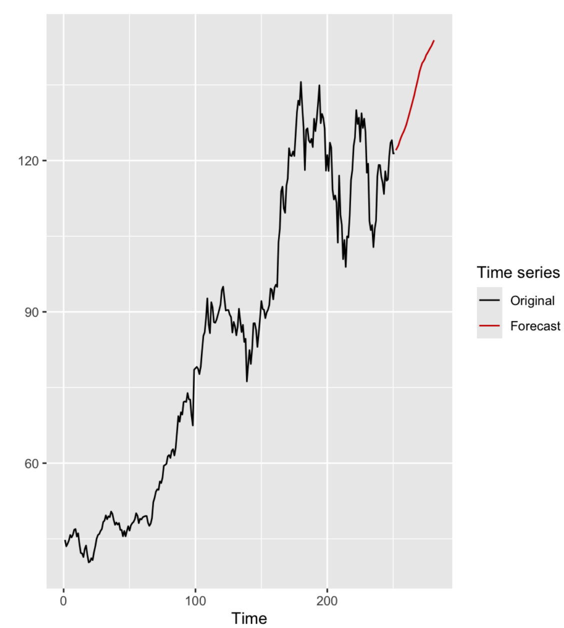
Figure 10 shows the AAPL K-NN forecast results using the same parameter as the SPY

Figure 10: AAPL K-NN



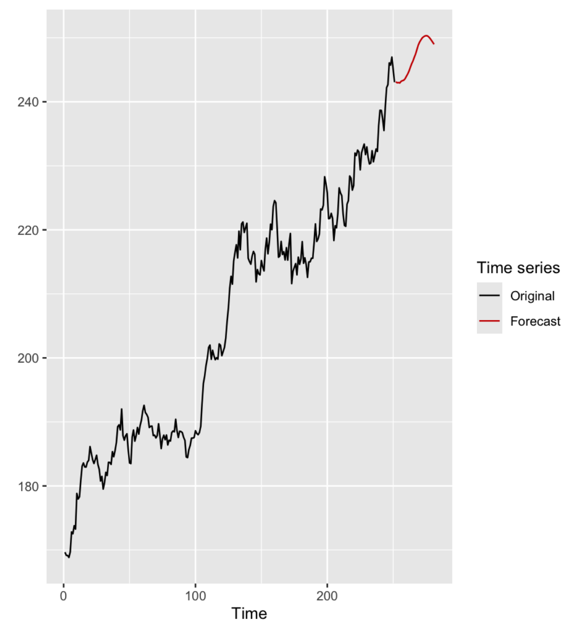
K-NN model. The rolling origins results as seen in Table 7 show a low number similar to the SPY results. We see the RMSE lower indicates a lower magnitude of the forecasting error. But we see that MAPE is almost double the SPY result indicating that the AAPL K-NN model shows a large deviation from the real data points.

Figure 11: NVDA K-NN



NVDA models as shown in Figure 11 indicate the forecast to spike, this may be due the the shocks experienced in the NVDA growth overall the year but is not indicative of its future growth. Our rolling origins are represented in Table 8, supporting that the model fit is not as great as the other two stock tickers. We see an RMSE almost twice as large and an MAE also almost twice as large which indicates a larger magnitude of forecasting error and further from actual values. This is also further supported by the MAPE which indicates about a 10% deviation from the actual value in our predictions.

Figure 12: GLD K-NN

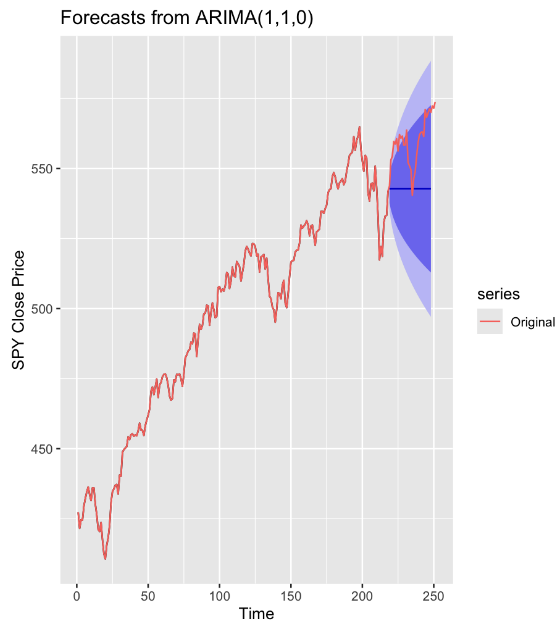


Our GLD K-NN model seems to be closer to our other models in terms of forecast graphs and rolling origins results. We see that RMSE and MAE are relatively low and MAPE is slightly higher than SPY but similar to AAPL. Even individually the graph in Figure 12 shows to be less variable then the NVDA results.

6 Robustness Check

In this section, we will discuss the robustness of our models. We will check the review test set to the training set to see the difference in the forecast and actual value. We will also take a deeper dive into thein K-NN rolling origins value and discuss the train and test set. We were not able to make a train and test set for the GARCH model within the given time frame and scope of the paper.

Figure 13: SPY Test Set ARIMA



The first comparison is the SPY ARIMA model in Figure 13. We see that the original is within the prediction of the ARIMA model. However, we do see that some of the predictions do not fit in the first confidence level but are still within our predictions.

Now let's compare this to the K-NN model, we do not have a confidence level for our graph but we can see that our overall prediction line is closer than to the ARIMA. This is also supported by our rolling origins table in Table 6, where outfit predictions deviate by just over 1% from the actual values. Both our other findings such as RMSE and MAE are low indicating that the fit is good and this is clearly shown in our graph.

The ARIMA model for AAPL seems to fit better than SPY as the model is within the dark blue area. The ARIMA model seems to be accurate when compared to the other stock tickers. We do know that the model is not perfect due to our Table 5 result but surprisingly the visual test does suggest the ARIMA is a good model.

Figure 14: SPY Test Set K-NN

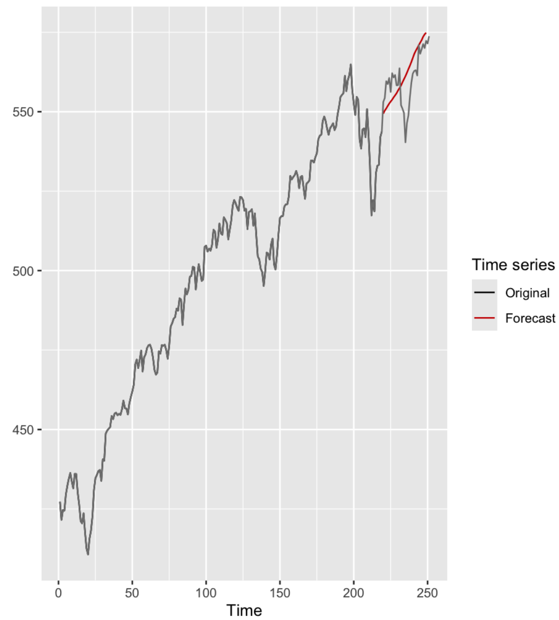
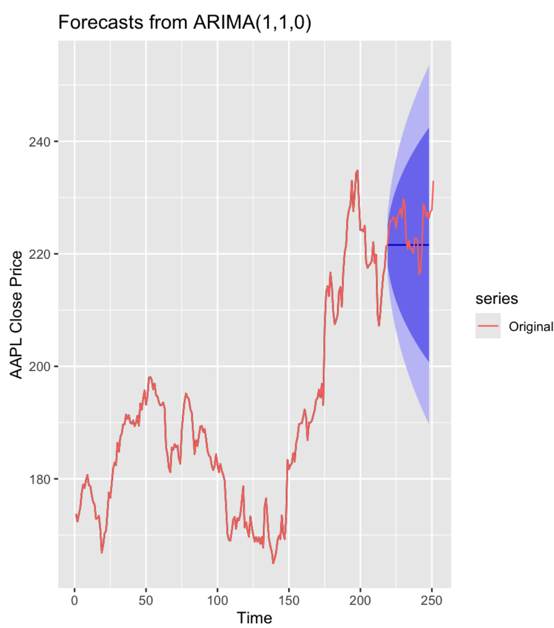
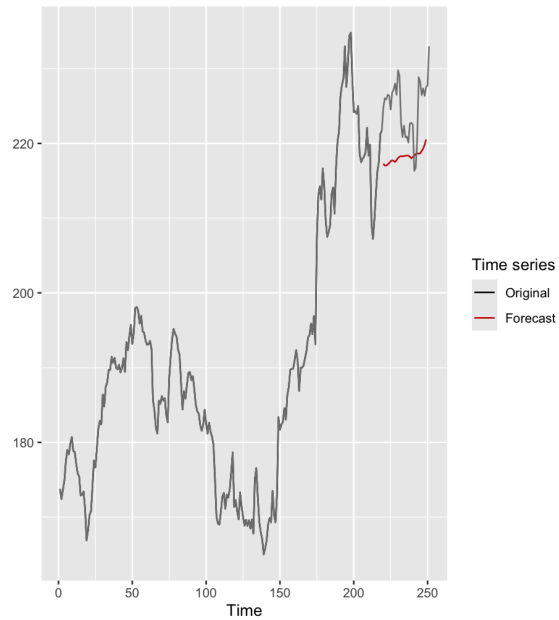


Figure 15: AAPL Test Set ARIMA



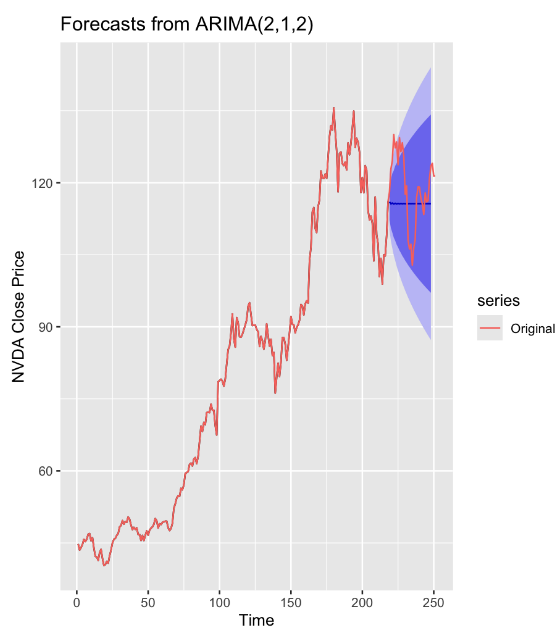
Going back to the K-NN model we see that fit is worse than the ARIMA model when accounting for the intervals in the ARIMA model. In addition, we do have the rolling average for this model and we do know it performs worse than then SPY test case as we know that the model has the overall worst result and this is indicated by eh visual in Figure 16 and the

Figure 16: AAPL Test Set K-NN



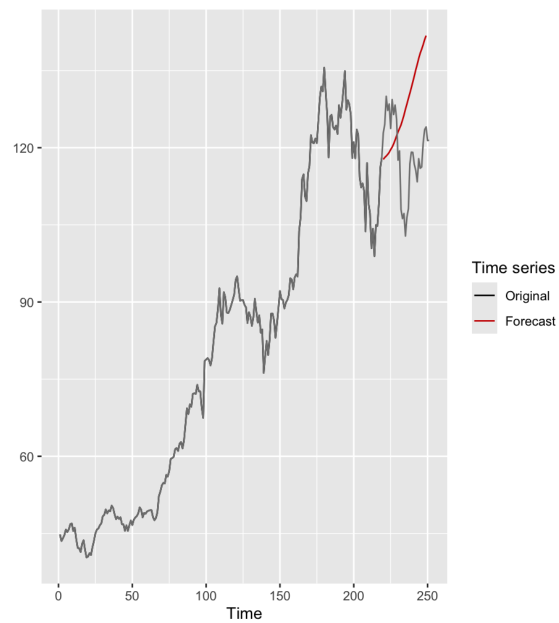
numerical figures in Table 7.

Figure 17: NVDA Test Set ARIMA



Again as briefly mentioned earlier we know that NVDA is the worst stock to predict because of the shock it experienced. This means that there is higher volatility in our results that the model may overestimate. This is clear in our ARIMA result and our Table 5 result indicating a bad fit compared to the SPY and AAPL models. We see that the original is not within the predicted ARIMA prediction in blue in Figure 17.

Figure 18: NVDA Test Set K-NN



Now compare that to the K-NN model which visually we see the forecast is directionally incorrect as we see that stock after shock starts to stabilize but due to the previous data we an increase in our forecast. This is also supported by Table 8 with about a 10% deviation from the actual value and this is again clear in our visual.

Going back to GLD which is a more stable ETF, we see that it performs better NVDA. We see that the original is within the forecasted intervals. But again none of the ARIMA models are statically significant, so we are going to have to take these results with some doubt in their validity. We can say so far our model is valid but further testing may be needed.

Compared to the more robust model we see the K-NN matches relatively close to the original visually. We can also see this in Table 9 where the MAPE is at around 2% and RMSE is low as mentioned before. In conclusion, we see the K-NN model and the ARIMA did a decent job visually but NVDA was a standout from the rest of the stock ticker.

Figure 19: GLD Test Set ARIMA

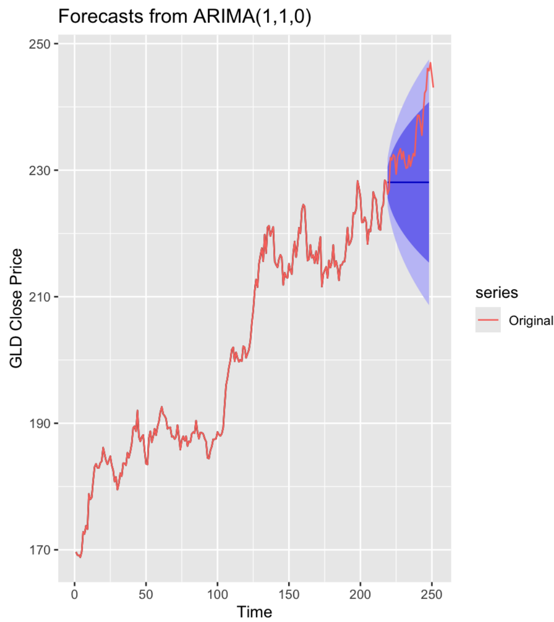
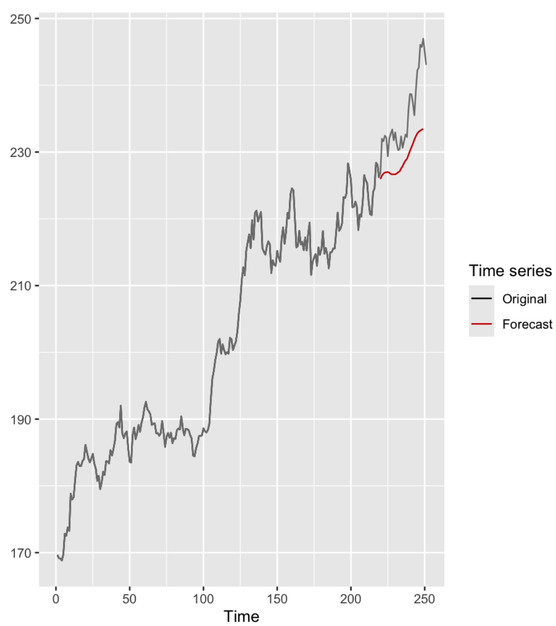


Figure 20: GLD Test Set K-NN



7 Conclusion

This research explores the application of ARIMA, GARCH, and K-NN models in forecasting stock prices, utilizing data from SPY, AAPL, NVDA, and GLD. Predicting stock movements is a challenge due to market complexities and volatility. Our research objective is to determine the optimal model for different types of stocks, referencing prior studies for model selections.

ARIMA models provide a baseline with a focus on autoregressive and moving average components. While the GARCH models capture volatility clustering common in financial time series. And lastly, the K-NN uses machine learning for non-parametric pattern recognition which leverages a more precise prediction based on clustering. Our result indicates that the ARIMA models had a limited statistical significance, especially for volatile stocks like NVDA, while GARCH models improved the fit by addressing shocks but still struggled with predictive accuracy. K-NN, however, emerged as the most robust model, particularly for stable assets like SPY and GLD, with low error rates and minimal deviations from actual values. However, all the models poorly predicted shock and adjusted for it in its prediction. The strength of the models, especially for K-NN is tied to the size of the dataset and forecasting horizon. We do not expect this model to be able to predict price tomorrow but to predict overall trends span months.

The robustness checks validate the K-NN model's predictive capability, supported by rolling origin error metrics which underscore its reliability. ARIMA and GARCH models provided insights into trends and volatility but lacked precision for actionable forecasting, particularly for stocks experiencing substantial market shocks. Overall, this study underscores the need for careful model selection based on stock characteristics and market conditions. Future research could enhance predictive accuracy by integrating hybrid models that leverage the strengths of ARIMA, GARCH, and K-NN while addressing their respective limitations.

8 Appendix

Figure 21: SPY

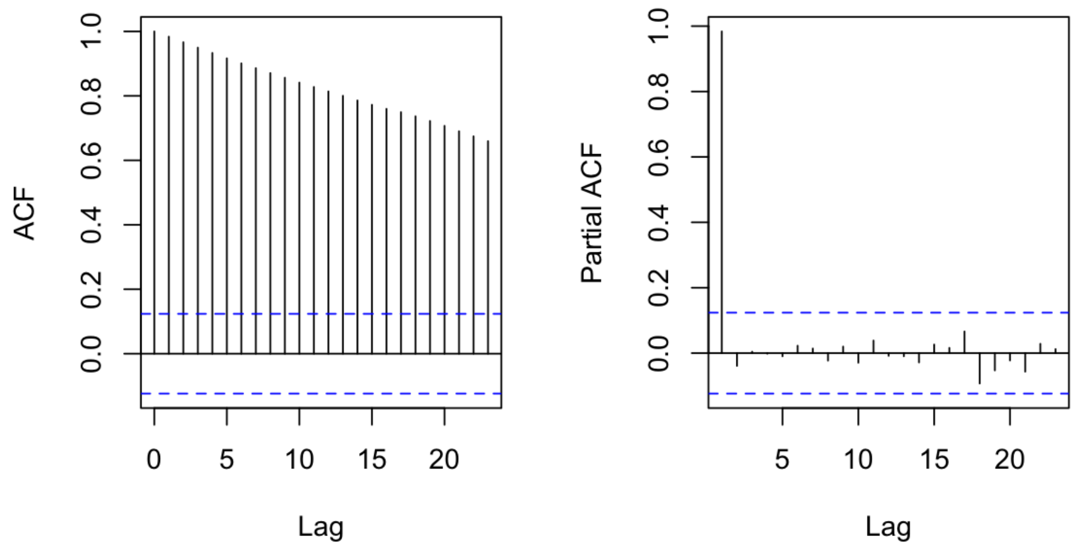


Figure 22: AAPL

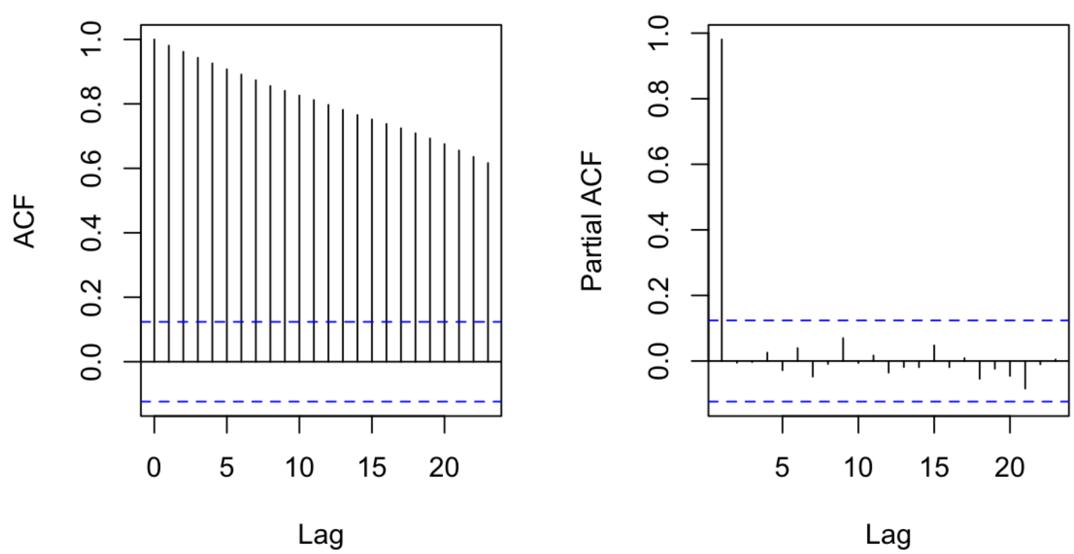


Figure 23: NVDA

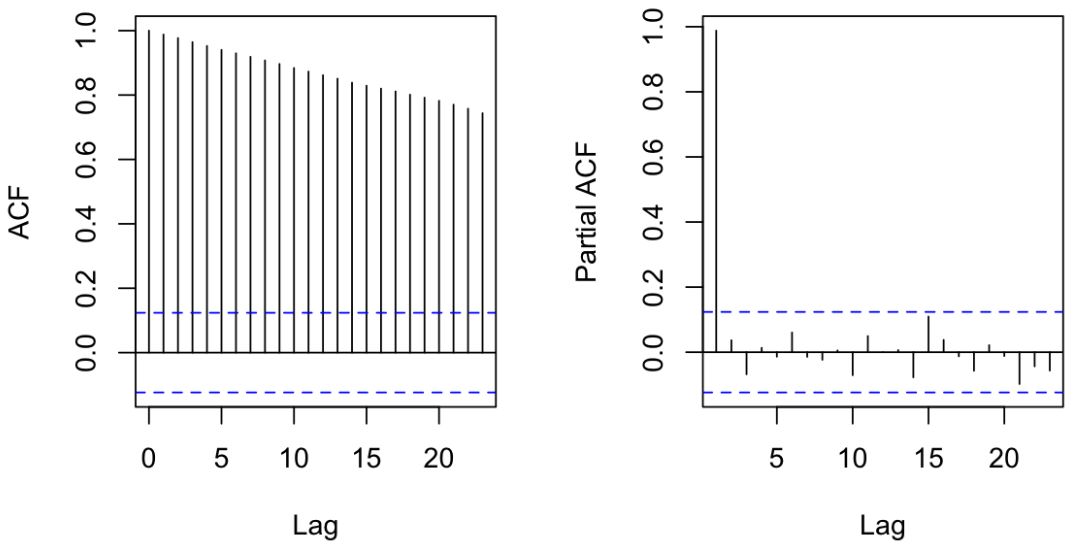


Figure 24: GLD

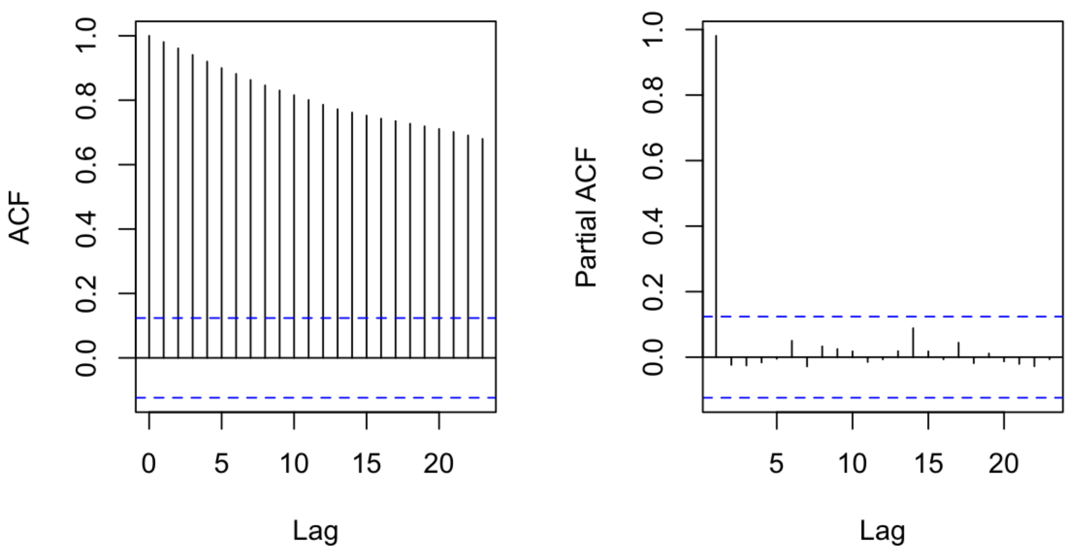


Figure 25: GARCH SPY

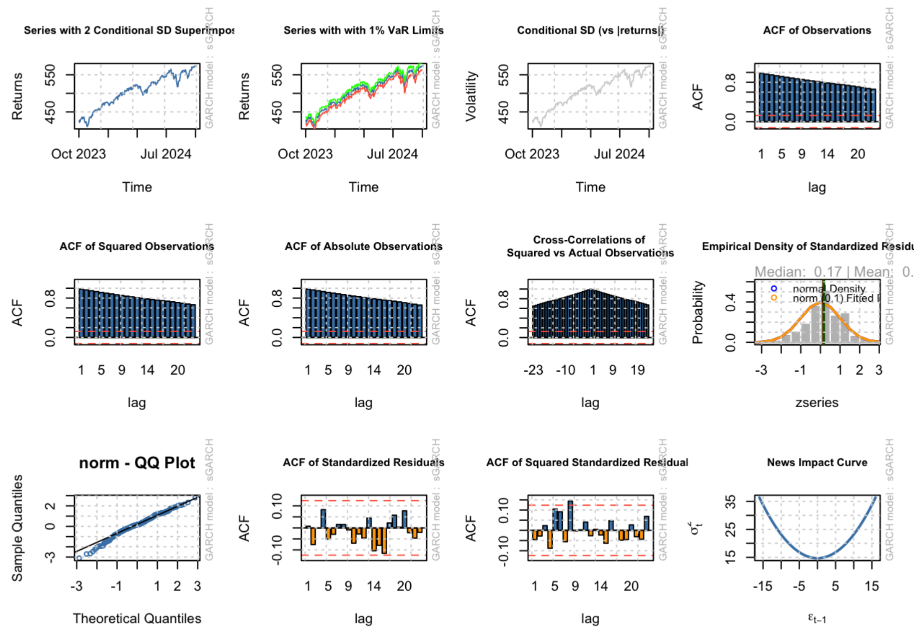


Figure 26: GARCH AAPL

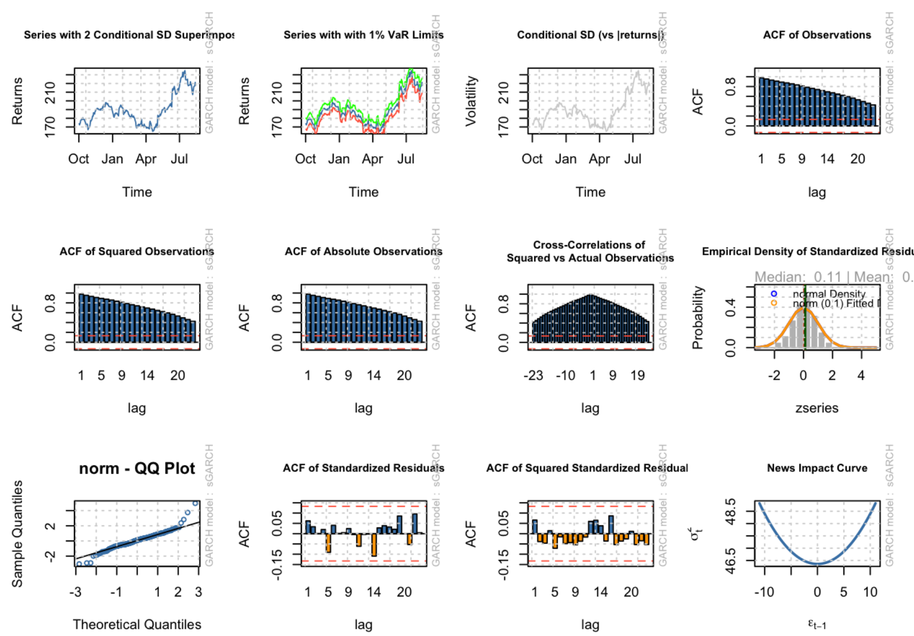


Figure 27: GARCH NVDA

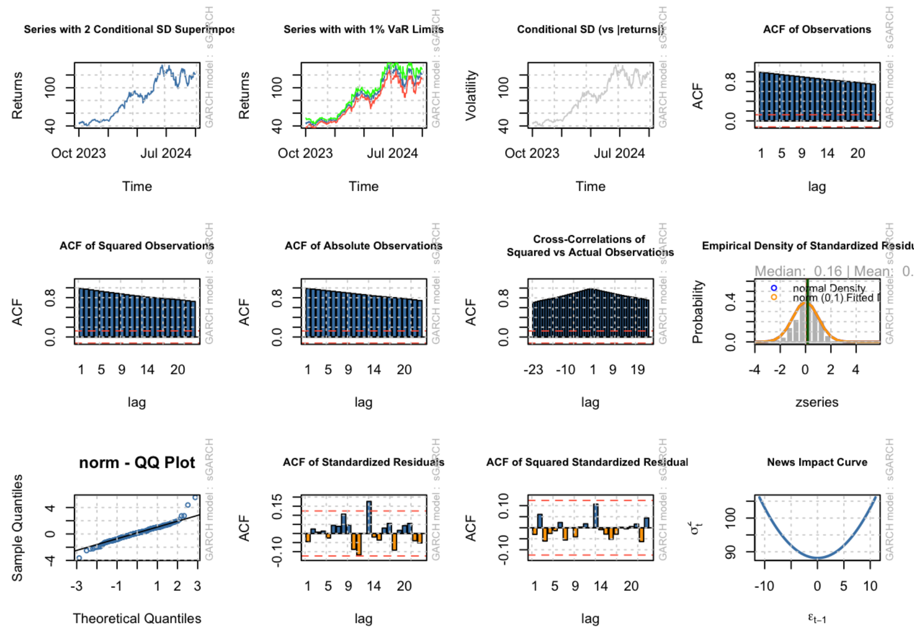


Figure 28: GARCH GLD

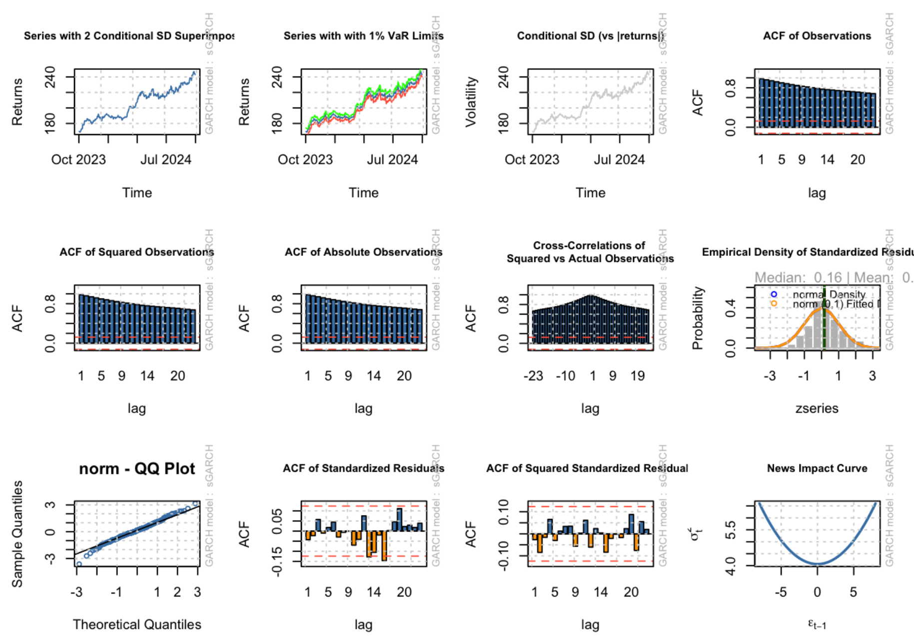


Table 5: ARIMA Model

	Dependent variable:			
	SPY	AAPL	NVDA	GLD
	(1)	(2)	(3)	(4)
ar1	0.041	0.084	−0.011	−0.045
	(0.063)	(0.063)	(0.553)	(0.063)
ar2			0.512	
			(0.422)	
ma1			−0.080	
			(0.567)	
ma2			−0.408	
			(0.392)	
Observations	250	250	250	250
Log Likelihood	−703.383	−606.834	−639.726	−518.799
σ^2	16.268	7.514	9.774	3.715
Akaike Inf. Crit.	1,410.767	1,217.668	1,289.451	1,041.597
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table 6: SPY K-NN Model Rolling Origin

RMSE	MAE	MAPE
7.636	5.726	1.028

Table 7: AAPL K-NN Model Rolling Origin

RMSE	MAE	MAPE
5.440	4.601	2.046

Table 8: NVDA K-NN Model Rolling Origin

RMSE	MAE	MAPE
13.726	11.642	10.125

Table 9: GLD K-NN Model Rolling Origin

RMSE	MAE	MAPE
6.102	5.103	2.112

References

Chatterjee, A., Bhowmick, H., and Sen, J. (2021). Stock price prediction using time series, econometric, machine learning, and deep learning models. In *2021 IEEE Mysore Sub Section International Conference (MysuruCon)*, pages 289–296.

Malagrino, L. S., Roman, N. T., and Monteiro, A. M. (2018). Forecasting stock market index daily direction: A bayesian network approach. *Expert Systems with Applications*, 105:11–22.

Ogundunmade TP, Adepoju AA, A. A. (2022). Exploratory data analysis and model performance of machine learning algorithms on economic data. *Modern Economy and Management*, 1:1–9.

Saha, S. and Bose, S. (2020). Forecasting of a time series (stock market) data in r. Accessed 22-11-2024.