Regression Analysis to Predict Apple Stock Volatility

Timothy Nguyen, Theresa Carotenuto, Francis Eshun,
Lekha Chunduri, Hannah Zhao

May 11, 2021

Stock Market Volatility

The focus for this report is the volatility of Apple stock returns. Volatility on the stock market is measured using the standard deviation to signal how tightly the price of a stock is grouped about the moving average. The industry tool for assessing market volatility is the CBOE Volatility Index (VIX). The VIX detects market volatility by calculating the implied volatility in the prices of a collection of put and call options on the S&P 500 Index. The S&P 500 is a weighted measurement stock market index of 500 of the largest companies listed on the United States stock exchange. A high VIX reading marks periods of higher stock market volatility, while low readings mark periods of lower volatility. Further, a high VIX signals higher risk in the market, making the VIX an effective tool for informing investor decisions. For this reason, estimating future market volatility is highly desirable.

Predicting Volatility

Economic factors at both regional and national levels, such as tax and interest rate policies, can significantly contribute to the directional change of the market and greatly influence volatility. A dissertation by Victor Perez of Pompeu Fabra University in Barcelona aimed to assess the volatility-predicting value of Twitter sentiment in conjunction with macroeconomic variables. It found that Twitter sentiment–namely public figures tweeting about a topic and thereby increasing network attention–explains a small part of market volatility that is outside the scope of macroeconomic factors. Perez found that his models performed adequately well on a normal day basis to forecast volatility. However, the trend prediction models did not perform as well in anticipating downtrends, and failed to predict unusually large peaks of volatility. This study suggests that an average day-to-day volatility can be well-modeled, but sudden peaks or downtrends remain difficult to anticipate.

A second publication from Embry-Riddle Aeronautical University titled *Using Multiple Linear Regression to Estimate Volatility in the Stock Market* [2] improves upon a prior regression model to predict the S&P 500 price. It acknowledges that the primary tool to estimate volatility today is the CBOE VIX, but asserts that this measure fails to show any warnings to a potential crash. The authors argue that the VIX is a measure that trails behind the market, returning only yesterday's numbers. They explore several relevant indicators – the consumer price index (CPI), producer price index (PPI), gross domestic product (GDP), money supply (MS), Unemployment Index (UI), VIX, and Fed Funds Rate (FFR) – and are able to pick out the most useful variables for an impressively simple and effective model. The final model includes only GDP, MS, and UI, all with very solid F statistics and an R^2 of 0.95. This model well approximates the movement of the market over a volatile two decades since 2000. It anticipates turning point of the market, known as the "market top" before a crash, and the "market bottom". The authors believe that this model can even be used to predict an alternate measure of volatility, which could then be inputted into other financial models. This improved measure is expected to perform better in financial models during volatile times.

Data Exploration

In this report, we examine several quantitative measurements that closely relate to Apple volatility over a two year time span. This data was gathered from FRED and Yahoo using the "quantmod" package for R. Specifically, it includes 503 measurements taken at the end of each trading day from January 1, 2018 to January 1, 2020. The first variable, the date, quantifies the number of days since January 1, 2018. Next, we have four measures taken from four different stocks: Apple, Google, Amazon, and the S&P 500. For each stock, the data set includes daily measurements of the closing price, the log return, the past 30 day volatility, and the VIX measurement, for a total of 16 predictor variables. This report will examine how these measures of stock performance correlate with the observed volatility of Apple stock returns over the next 30 days. We construct, analyze, and refine several linear regression models that assess how different combinations of these predictor variables can best predict Apple stock volatility.

Single Variable Linearity

To begin this exploration, we first examine the relationship between our response variable–Apple Vol–and each predictor individually. We construct box plots to show the distribution of each variable, and scatter plots to show the relationship between the response and each predictor. From this, we identify potential outliers and assess the possibility of a linear relationship between Apple volatility and the other variables.

We found that variables within the same measurement category all demonstrated similar relationships with the response. For example, the four variables quantifying the log returns of Apple, Google, Amazon, and the S&P 500 all had very symmetrical box plots with some outliers on both ends of the range. The scatter plots of return variables indicated almost no linear relationship with Apple volatility. These distributions are what we would expect to see from this measure; returns are generally clustered around zero because they represent an incremental change from yesterday's return.

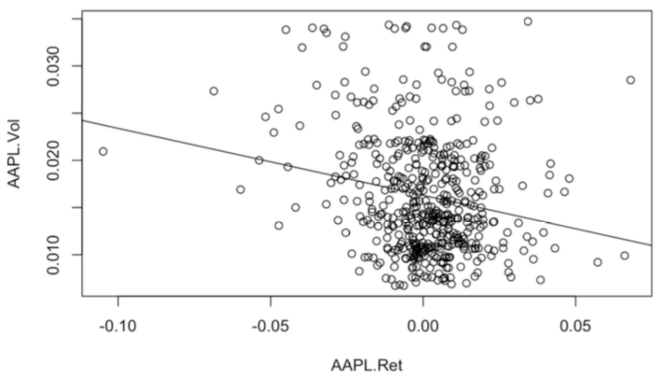

Figure 1: Scatter plot: Apple Volatility versus Apple Returns

The Past Volatility variables had a skewed distribution with some outliers at the extreme volatility values. The scatter plots suggest a very slight linear relationship with Apple volatility.

Next, the distributions of the Closing Price variables were moderately symmetrical, but a handful of outliers on either end of the range could be expected. The scatter plots also revealed a very slight negative linear trend between closing prices and Apple volatility. Finally, the VIX variables demonstrated the strongest linear relationships with Apple volatility. The scatter plots showed the most linearity compared to scatter plots of the Returns, Closing Prices, or even Past Volatility. The box plots all had a skewed distribution with outliers at the high end of the range of values. The Apple Volatility box plot itself also has the same skewed distribution with similar outliers.

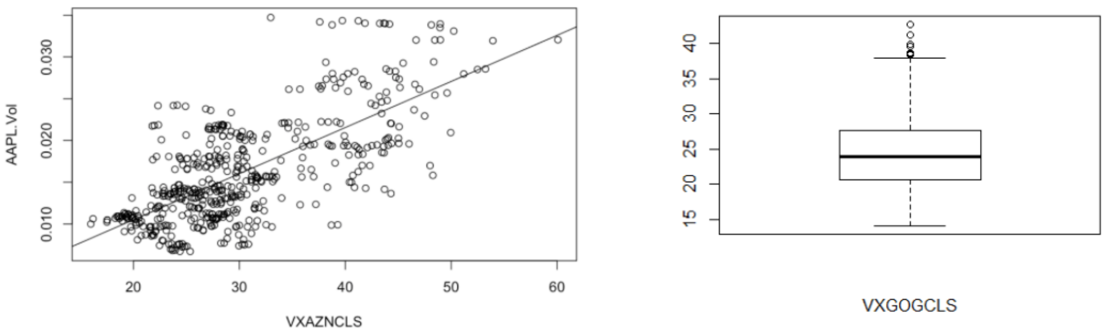

Figure 2: VIX Variable Plots against Apple Volatility

We also checked for significant correlations between predictor variables by creating a Pearson correlation matrix. Evidently, variables within the same general category–past volatility, closing price, log return, and VIX–are highly correlated. From one category to the next, it's interesting to see which variables are more closely correlated. For example, the past volatility variables have a slight positive correlation with the VIX variables, and a moderate negative correlation with the closing prices. The returns have generally negligible correlation with any other variables. This matrix also supports that Apple Vol has the strongest positive correlation with the VIX variables.

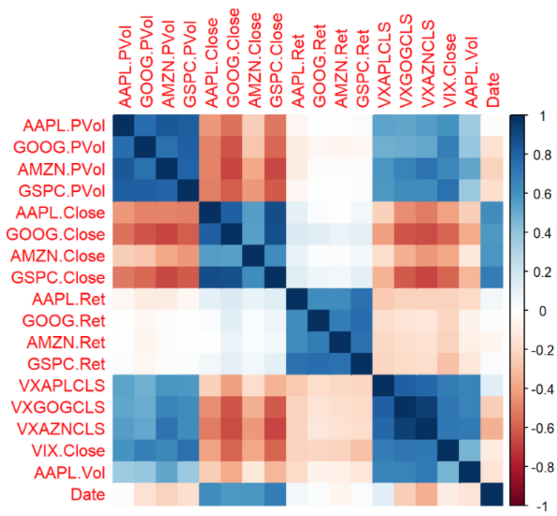

Figure 3: Correlation Matrix

One final measure in our single variable exploration was to construct a 95% confidence interval for each variable to assess whether there was a statistically significant linear relation with the response. We found that several predictors had an extremely small Beta, but the interval still did not include zero. The VIX variables had confidence intervals that did not include zero, but that hovered around 0.0005 to 0.0008. Close variables found even smaller Beta values, but they were negative. The sole predictor that was ruled out as having no linear relation with Apple Vol was the Google Return, with a 95% Confidence Interval of [-0.0617, 0.0067].

Linear Regression Analysis

Moving on to some more in depth analysis, we constructed linear regression models between Apple volatility and each predictor and conducted a residual analysis to determine if this linear model was the appropriate fit for the data. We checked for outliers, linearity, constant variance, and normality and independence of error terms.

Several Q-Q plots were slightly concave up, as is shown below. This graph is the Q-Q plot derived from the Apple Vol and Apple closing price linear regression model. This trend was present in many Q-Q plots, so it was difficult to determine the significance of this non-normality from one regression to the next at this stage.

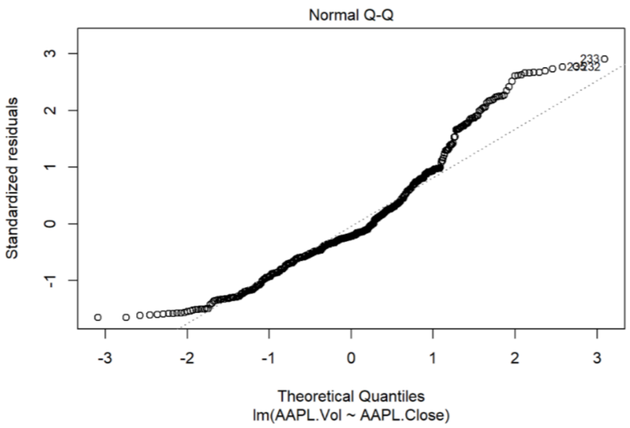

Figure 4: Normal Q-Q Plot

The R^2 term in most models was less than 0.2, which suggests that some transformations to the linear models would be necessary. The VIX variables had the strongest linear relationship with Apple Vol. The residual plots looked like the one shown below, and we observed R^2 values as high as 0.4941.

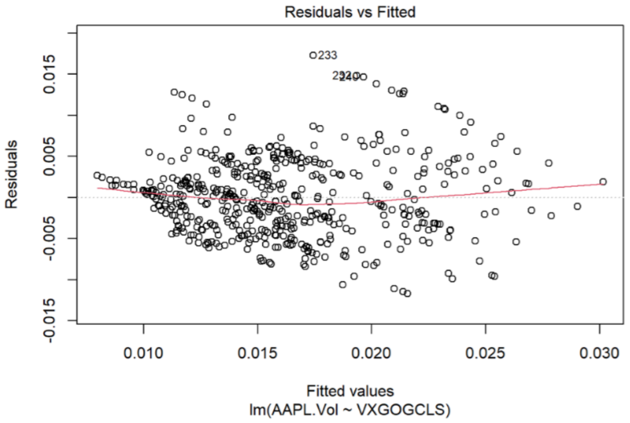

Figure 5: Residual Plot with VIX Predictor

We used these diagnostics to inform potential transformations of any variable, and tested several transformations to see if they were a better fit for the data. One of the more interesting transformations observed was the Apple Volatility vs VIX variables. We found that some logarithmic transformations on Apple volatility created a more normal residual for our Q-Q Plots considering extreme values. Apple volatility vs VIX Google presented as follows:

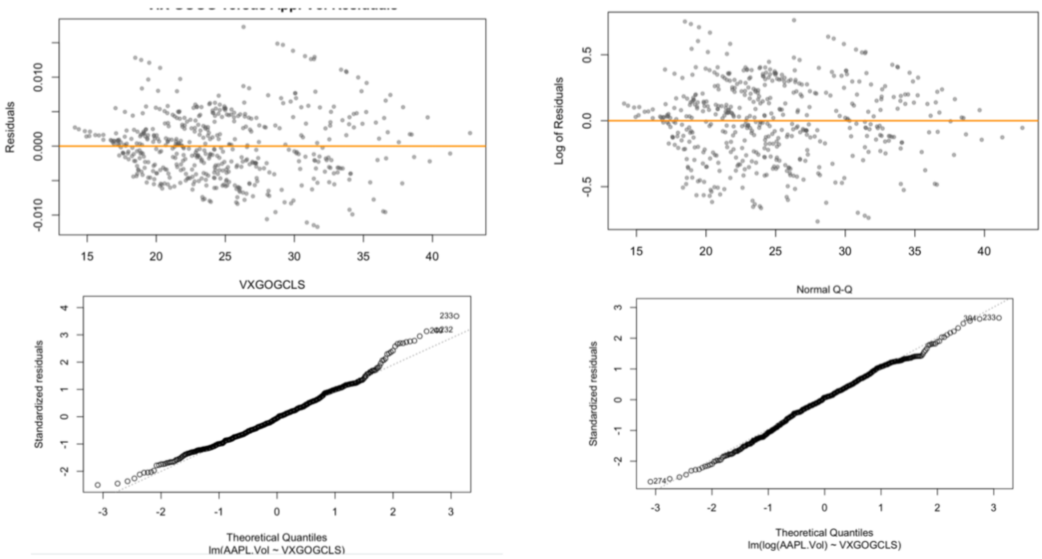

Figure 6: Original Plots (left); Transformed Plots (right)

After this preliminary single variable model investigation, we have a few important takeaways to note. According to our Pearson coefficients, we know that several variables have high correlations with predictors from the same measurement category. Our next models will likely include interaction terms of some combination of these variables. At this stage, there are no significant outliers that need to be dealt with. We found some transformations to be marginally helpful in improving linearity, but we must continue to more complex models if we want to accurately predict Apple volatility.

Model Selection

R functions

We utilized some packages to help generate our models. For plotting our models and residuals, we utilized ggplot as well as the default R package. We also used the olsrr and regsubsets packages to regress our models and provide us with useful metrics like AIC, SBC, and RMSE.

Forward Stepwise Regression

When we did forward stepwise regression, we got a subset of 15 variables out of the 17 possible variables. The variables added were VXAZNCLS, AMZN.Close, VXAPLCLS, Date, GSPC.Close, AMZN.PVol, GSPC.PVol, GOOG.PVol, GOOG.Ret, GSPC.Ret, AAPL.Close, AAPL.PVol, GOOG.Close, VXGOGCLS, and AMZN.Ret.

Backwards Elimination

When we did backward stepwise regression based on p-value, the model removes two variables: Vix.Close and AAPL.Ret.

Backwards Stepwise Regression

When we did stepwise regression, we were able to get a subset of nine variables. The variables that we ended up with was: VXAZNCLS, AMZN.Close, VXAPLCLS, Date, GOOG.PVol, GSPC.PVol, GSPC.Close, AMZN.PVol, and GOOG.Ret.

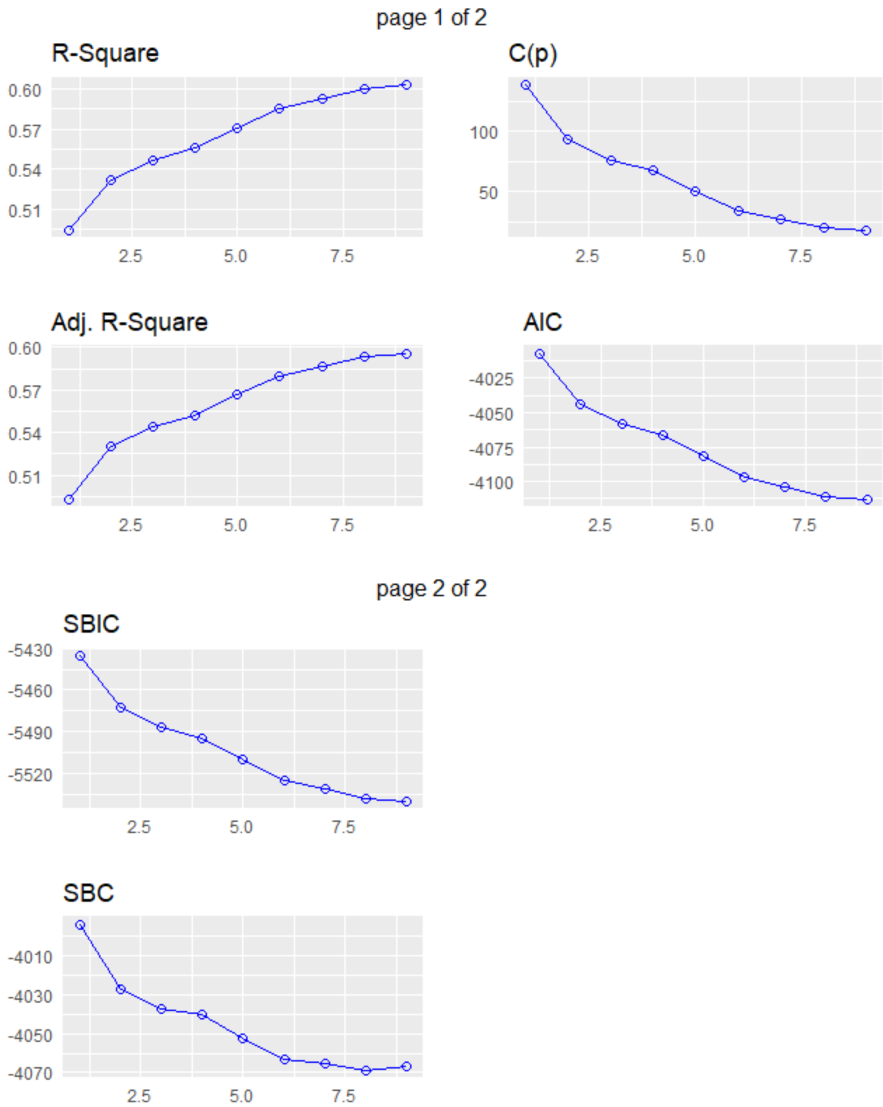

Figure 7: Stepwise Regression Metrics

Collinearity Diagnostics

Using the collinearity diagnostic function from the olsrr R package, we calculated the Variance Inflation Factor of the predictors in this initial subset. We did not find any variable to have a VIF greater than 10. The most concerning variable was VXAZNCLS, with a VIF of 7.95. The VIX on Apple and the GSPC Close also had a VIF of 5.79 and 5.16 respectively. These predictors warrant further investigation to determine if the level of collinearity is problematic.

Model Refinement: higher order and interaction terms

We expected our model to require some higher order terms and interaction terms to remedy to high correlations between predictors. The correlation matrix of the nine initial variables in our model revealed that the most noteworthy correlations exist between variables within the same measurement category. We would have to test for an interaction term between the two VIX variables, the two Close variables, and the three Past Volatility variables. One by one, we investigated a pair of predictors separately against AAPL.Vol. We centered the variables in question, and added the interaction term to the original model. Using F-tests we could confirm whether the interaction term was a significant addition to the model or if it could be dropped.

Initially, each interaction term individually had an F statistic that required it remain in the model. To investigate further, we tested models with one interaction term against a model with two interaction terms. Eventually, through repeated F tests and confidence intervals, we were able to refine the model. The hypothesis tests for Beta found that we could drop the interaction term between the VIX variables and the interaction term between the Close variables. Additionally, we were able to drop the interaction term between Google Past Vol and GSPC Past Vol when other Past Vol interaction terms were included.

At this point, our model includes VXAZNCLS, AMZN.Close, VXAPLCLS, Date, GSPC.PVol, GSPC.Close, AMZN.PVol, GOOG.Ret, and the interaction term amznpvol*googpvol. The current multiple R-squared value is 0.6234.

Next, we test each variable in this model to see if it requires the inclusion of a higher order term. We added the square of a term to the model one at a time and conducted F tests to determine if the higher order term could be dropped. The higher order terms that improved the fit of the model significantly were AMZN.PVol^2, GOOG.PVol^2, Date^2, and GSPC.Close^2.

Incorporating these results, we arrived at the following model: $-0.08878+0.0004351(\text{VXAPLCLS})$ $+0.000009089(\text{Date})+0.0003130(\text{VXAZNCLS})-0.00002999(\text{GSPC.Close})+0.0003130(\text{VZAZNCLS})$ $+ 0.2787(\text{GSPC.PVol}) - 125.5(\text{GOOG.PVol} * \text{AMZN.PVol}) + 60.19(\text{AMZN.PVol}^2)$ $- 0.00000003747(\text{Date}^2) + 56.36(\text{GOOG.PVol}^2)$.

Although several variables have a β - coefficient very close to zero, hypothesis tests for each Beta concluded that the confidence intervals did not include zero, and thus the variable could not be dropped.

Outliers and Influential Points

With our final model, we observe an $R^2 = 0.712$. We then check for outliers and influential points to see if any of the data points could be removed to further improve our model's fit.

From observing our Deleted studentized residuals, we observed that none of the data points in our model stood out as an outlier. Upon observing our DFFITS and our DFBETA's we observed many influential points in our data. This was more clear in our Cooks distance plots as it was easier to see overall how our response variable fared against all variables in a single direction as opposed to the opposing directions in our DFFITS plots. From observing our Cooks distance plots (Figure 5), we can see that none of the data points are above a value of 0.05, which means that its reasonable to justify that there are no overly influential points in our model. We also observed that the large majority of influential terms in our DFBETAS were the same data points for our influential outlying points in our Cook distance plot. For the purpose to testing, we remove the influential points that are ≤ 0.02 see observe how much influence they can indeed have on our model. Our threshold of ≤ 0.02 was chosen in order to account for the influential DFBETA points and, see how it affects our model's R^2 value.

From our analysis of influential points, we are able to observe 4 very influential points in our model. Removing these 4 points gave us a newer $R^2 = 0.7248$. This is a better fit for our model. Doing model verification we can check whether our model is valid or overfitted.

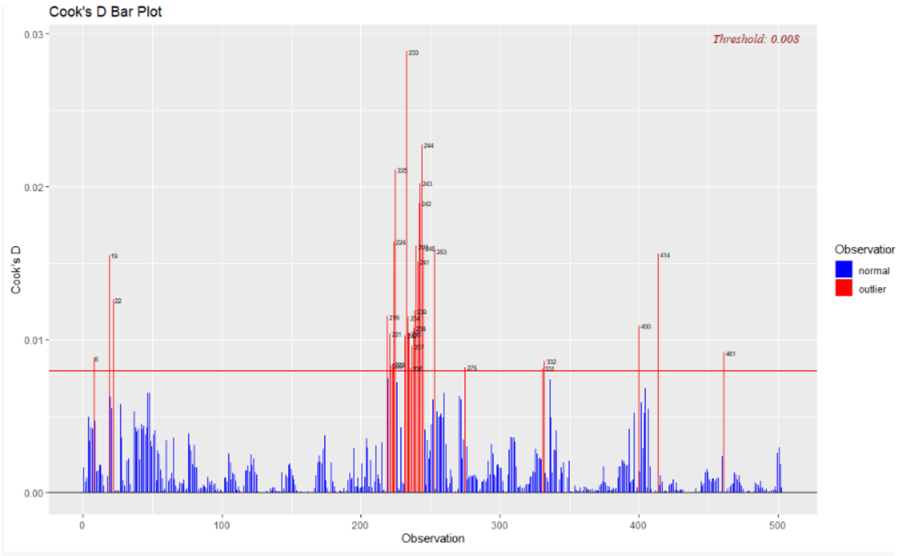

Figure 8: Model Cooks Distance Plots

Model Validation

To validate our data, we ended up doing a few things. We utilized a repeated k-fold cross validation to validate the information since our data sample is limited. We first shuffled the dataset randomly. Then, we split the dataset into k groups, which in this case is 10. For each group we basically train nine out of the first ten parts of the dataset. Then, we test the other part by evaluating it and fit the model accordingly. We train this ten times. When we validate the model, we end up getting a $R^2 = 0.7188775$. We also had a RMSE score of 0.003298721 and MAE score of 0.00258. This was a significant improvement to the original model, where the best model was hovering with a $R^2 = 0.61$.

We also validated the model by looking at the F-score for each portion and did a confidence interval test. The F-score for each section passed the $\alpha = 0.05$ test and for almost all but one, it passed $\alpha = 0.01$.

```
Response: AAPL.Vol
                          Df    Sum Sq   Mean Sq  F value    Pr(>F)    
VXAPLCLS                   1 0.0083120 0.0083120 769.8417 < 2.2e-16 ***
I(GOOG.PVol * AMZN.PVol)   1 0.0005165 0.0005165  47.8347 1.457e-11 ***
Date                       1 0.0006568 0.0006568  60.8344 3.781e-14 ***
Date2                      1 0.0003383 0.0003383  31.3285 3.635e-08 ***
AMZN.PVol2                 1 0.0003611 0.0003611  33.4477 1.307e-08 ***
GOOG.PVol2                 1 0.0015678 0.0015678 145.2100 < 2.2e-16 ***
GSPC.Close                 1 0.0013989 0.0013989 129.5662 < 2.2e-16 ***
VXAZNCLS                   1 0.0004401 0.0004401  40.7577 4.009e-10 ***
GSPC.PVol                  1 0.0000906 0.0000906   8.3918  0.003938 ** 
Residuals                489 0.0052797 0.0000108                      
```

Figure 9: F-table

For the CI interval, none of the variables contains 0 so all of these variables are significant.

	2.5 percent	97.5 percent
(Intercept)	-1.037691×10^{-1}	-7.378473×10^{-2}
VXAPLCLS	3.156397×10^{-4}	5.545652×10^{-4}
I(GOOG.PVol * AMZN.PVol)	-1.465097×10^{2}	-1.045156×10^{2}
Date	2.529265×10^{-6}	1.564857×10^{-5}
Date2	-4.783646×10^{-8}	-2.710561×10^{-8}
AMZN.PVol2	5.066039×10^{1}	6.971802×10^{1}
GOOG.PVol2	4.758279×10^{1}	6.513543×10^{1}
GSPC.Close	2.498105×10^{-5}	3.500577×10^{-5}
VXAZNCLS	2.135104×10^{-4}	4.125441×10^{-4}
GSPC.PVol	8.967355×10^{-2}	4.677550×10^{-1}

1 Summary

In conclusion, our final model that we came up with was:

$$\begin{aligned}
-0.08878 + 0.0004351(VXAPLCLS) + 0.000009089(Date) - 0.00002999(GSPC.Close) \\
+0.0003130(VZAZNCLS) + 0.2787(GSPC.PVol) \\
-125.5(GOOG.PVol * AMZN.PVol) + 60.19e(AMZN.PVol^2) - 0.00000003747(Date^2) \\
+56.36e(GOOG.PVol^2)
\end{aligned}$$

Essentially, it's a combination of eight predictor variables taken from our original dataset, with one interaction term comprising two of them (GOOG.PVol and AMZN.PVol), and three quadratic terms coresponding to three of them, Date, AMZN.PVol and GOOG.PVol. Summarizing what were the steps carried out for obtaining this final model:

- As part of the **initial data exploration**, we plotted the correlation matrix to look at all the relations between the predictor variables and the response variable and amongst each other, we plotted the residuals and examined them, and applied some transformations on our response variable and on predictor variables; in the case of the predictor variables, it was to make singular predictor variables normal, but realizing this actually made our model worse, we neglected them.
- For the **model selection**, we first did the step-wise regression, both forward and backward, and backward elimination, from which we were able to select nine variables to go forward with. We then carried out tests for higher order terms and interaction terms, and incorporated the ones into our model which were relevant and didn't risk overfitting.
- For **model diagnostics**, we looked at the collinearity between the variables and searched for any outliers and influential points.
- Finally, for **model validation**, we carried out cross validation and obtained low RMSE and MAE values, ensuring our model is accurate.

2 Conclusion

Points about interpretation of results: - apple's stock volatility can be predicted for the most part through observing the nature of stocks of other associated technological companies such as Amazon, Google, SP 500, over that of apple itself - VIX variables were included and Returns were not; this is reasonable, as CBOE Volatility Index indicates the market's expectations for volatility of companies over the coming thirty days, and it's in real time. This is already a given verifiable measure found by reliable sources, and it's well-known, so it holds power to impact the decisions of stock buyers, over knowing the value of the stock return rate on that particular day. - because our model has both negative and positive beta values, this shows our predictor variables impose both decreasing and increasing effects on the Apple volatility - from observing how stocks have deviated in the past, it can help us forecast how stock volatility will be in the future, this reflects in our mode, since three out of 4 .PVol variables are included.

(kill me)

With this, our closing statement is as follows: We can say that, in light of how the Apple stock returns have deviated during a two-year time span from 2018 - 2020, the standard deviation of Apple's log returns in the coming 30 days can be forecast predominantly from analyzing the log returns from the previous 30 days of Amazon, Google, and SP 500 separately, the particular date within the two-year period, the closing prices of SP 500, and VIX (CBOE Equities) on Apple and Amazon separately.

References