Timothy Nguyen
Statistics 525: Group 6
Reflection: A Regression Model To Predict Apple Stock Volatility

In this area of research, there's important findings that offer insights into how volatility occurs based on trends in the market. We found that the model predicting apple volatility initially selected for VXAZNCLS, AMZN.Close, VXAPLCLS, Date, GSPC.Close, AMZN.PVol, GOOG.PVol, GSPC.PVol, and GOOG.Ret in our backwards stepwise regression model. Most of these variables were chosen because these variables are highly correlated to Apple Volatility. When doing a single-variable analysis with a correlation matrix, the most negatively correlated variables are GOOG.Close and GSPC.Close. The most positively correlated variables are AMZN.PVol, VXAPLCLS, VXGOOGCLS, VIXAZNCLS, and VIX.Close. However, there are certain variables that weren't necessarily correlated highly with APPL.Vol were in the actual model such as the Date variable. However, there are really strong correlations from variables in the model such as AMZN.Close and GSPC.Close. There are some variables from our initial model that was a bit strange to see be added such as the Google return value. However, it was removed in our later models as it failed the F-score component in our model validation portion of the report.

Additionally, when looking at daily log returns and closing prices and VIX on Google, Apple, Amazon, and the whole S&P 500, the main predictor over the past volatility looked to be a major predictor in modeling for apple volatility. This could be explained by the fact that how Apple generally does in the market is also affected by its competitors, which are Google and Amazon. They may depend on one another. For example, Apple's cloud infrastructure is hugely dependent on Amazon Web Services, which is a cloud computing platform that provides services and APIs from individuals to corporations. Currently, Apple pays around $30 million a month for Amazon Web Services in order to operate its 1.5 billion active devices. Also, nearly half of Google's search traffic comes from Apple's devices. It could explain that even if these companies can be competitive with one another, they are in some way dependent on one another. When looking at the model, the past volatility of Amazon and Google were relatively positive with a coefficient of around 0.3. The coefficient of these terms in our final model was 60.19 and 61.83 respectively and had an interaction term that was -125.5. Because these coefficients were some of the larger ones we found in the model, it could indicate that these measures had the largest impact on apple volatility. It could also mean that apple's volatility can depend on how volatile the google and amazon past 30 day volatility is.

It is also interesting that in our final model, there are very small coefficients in the model. The closing, vix, and date indicators show that there's a really small coefficient of almost zero. However, when running our 95% confidence interval tests, none of those variables contained zero. Additionally, the F-tests calculated from the anova table generated from the model showed that each variable passed the alpha = 0.01 test for all variables involved in the model. What could be resulting from these results is that there's a very narrow range of values for these variables that won't change the predicted apple volatility. It's most likely that the model has a lot of small values that might create some noise or background to the actual model. For example, for VIX and volatility, traders typically want to sell when the volatility is higher and buy when the volatility is lower. There could be a small, but significant relationship because there's only certain times in the economic cycle when these stocks can be really volatile or really stable based on how the overall market is doing.

There were many surprising components that occurred when looking at the results of our model. The date variable was included into our final model where date was a positive coefficient and its higher order squared term was negative. There could have been some issues where the performance of how apple volatility was affected by surrounding days' outputs.

A way to improve upon our current model is to possibly explore more into centering variables and possibly add them one by one to the model more formally for the next time. Additionally, we initially had a model that included higher order terms ranging from the 4th to 6th term. It might be interesting to explore this relationship later down the road and see if they end up giving us good results or make the model worse.