Whose Correlation Is It Anyway?: Forecasting S&P 500 Realized Volatility Using Macro Correlations

Abstract: This paper will present the case for using the volatility of macro variables which are commonly sited for use in Financial Conditions Indicators and Indicators of Financial Stability to forecast realized volatility of the Equity market, specifically the S&P 500, which we will proxy by using the ES1 futures contract. We will use these variables to explain realized 5-day volatility in the S&P 500 Index and then attempt to use them for forecasting the next 5 days realized volatility using a Linear Regression to help analyze and demonstrate the viability of our variables and then the popular XGBoost Regression model to capture more complex relationships. The final model will be rolled forward throughout the period of 2015-2022 demonstrating purely out of sample performance. We will also specifically examine the performance of the models during the Covid-19 volatility spike in 2020.

1. Rationalization of Variables

The variables used will be based on the time series of prices for various financial instruments. These financial instruments are used in the formulation of the Financial Stress Index, "https://www.financialresearch.gov/financial-stress-index/files/indicators/index.html" , and in the Goldman Sachs Financial Conditions Index, "https://www.goldmansachs.com/insights/pages/case-for-financial-conditions-index.html" . After iterating through linear regressions using a larger set of variable chosen for intuitive reason and removing those whose p-values showed that they were non-informative I arrived at a final set of variables. The variables chosen are present in both indexes and all represent heuristically important macro variables. The time series for the variables starts in 2003, so our analysis will begin at that time which gives us enough data to examine the robustness of the model. The seven financial time series chosen are given and briefly described in the following table:

1.1 Feature Engineering

The engineered features used in the models will be the Exponentially Weighted Ten-Day Moving Average Volatility and the rolling correlations of these assets' returns with the ES1 returns for both the last 252 days and the last 5 days. The VIX index is not transformed. I also tested all of these features for stationarity. All had p-values very close to zero for the Augmented Dickey-Fuller test so they are not found to be non-stationary. The next test is so that we can use these features in a regression which will be the first step of modeling analysis.

1.2 Data Analysis

Multicollinearity will cause a linear regression to give faulty results and is also undesirable in many statistical learning techniques. Correlation with the target variable is often desirable as it shows there may be some predictive power. Below is a correlation matrix of our variables plus the two year swap index volatility and the ES1 Vol of Vol. These features were not found to correlate well or be predictive in linear regression or more advanced ML models so they were excluded from future analysis.

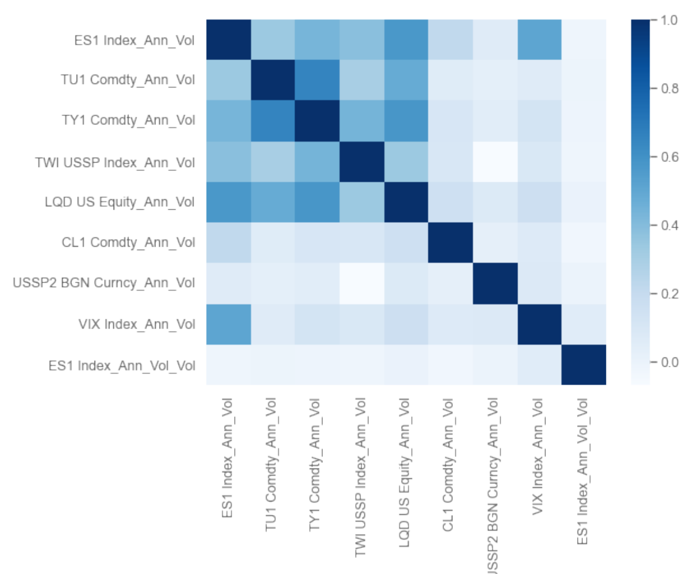

Figure 1. Correlations Between Asset Volatilities

Exhibit. 1. Macro Variables

Variable	Description
ES1	The active futures contract for the S&P 500.
TU1	The active futures contract for the 2 Year Treasury Note.
TY1	The active futures contract for the 10 Year Treasury Note.
TWI USSP	The trade weighted dollar index, representing dollar strength against other G-10 countries.
LQD	LQD ETF which represents US Corporate Bonds.
CL1	The active futures contract for crude oil.
VIX Index	The implied volatility of the S&P 500.

We can see from the correlation heatmap that none of our variables are overly correlated while they do all have some level of correlation. The hope for our model is that using both the volatilities and the correlations of these assets we can identify which asset volatilites are influencing stock market volatilities at different times as these correlations rise and fall.

The exponentially weighted volatility contains recent information on the volatility of each of these assets and is weighted toward the most recent observations. Volatility has well documented serial correlation which makes the EWMA a forward-looking indicator. Changes in asset correlations also indicate changes in the market as in times of turmoil steady correlations will often change. This is why I include the weekly, monthly, and yearly correlations of these time series with the ES1 time series for the final model.

Visualizing the standardized volatilities of our chosen features vs the ES1 Realized volatility is also a useful tool to rationalize picking these variables. The volatilities are standardized for this visualization by taking the rolling one-year z score at every data point. Otherwise, they are not comparable as the scales of these volatilites vary greatly. In the below graphs, we can see on different vol spikes one asset may mirror the ES1 while others do not.

Figure 2. Rolling Z-scores of 5-day realized volatility of Variables.

The rolling correlation features need to be similarly visualized in order to gain some intuition as to how these various time-scaled correlations can help us forecast volatility. to do that we examine plots of the 5, 21, and 252-day realized volatilities of our variables along with the 5-day realized volatility of the ES1 future. In the below graphs, we can clearly see the divergence of the 5 and 21-day volatilities from the 252-day in times of spiking volatility. We do not see every asset exhibiting this behavior for every volatility spike, but usually at least one does. This is encouraging, a tree-based model such as XGBoost which benefits from bagging and boosting and can pick up on highly non-linear functions should do well with these relationships.

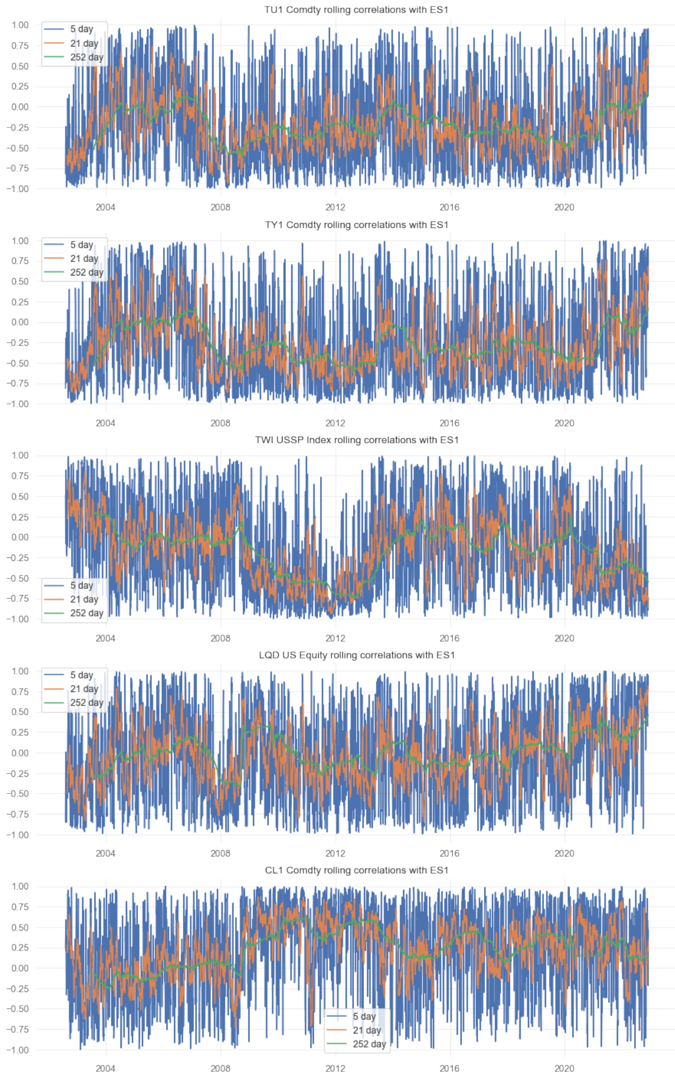

Figure 3. Rolling Correlations of various spans for macro variables.

2. Modeling Methodology

The data analysis gave reason to believe these variables will have some predictive power over the volatility of the S&P 500. The next step is to examine whether we can find linear relationships between these volatilities in linear regression and then if we can use these volatilities for forecasting the week ahead realized volatility of the S&P 500. Through this analysis, the 2-year swap index and the vol of vol variables were dropped as they exhibited no predictive power in regression or in XGBoost.

2.0.1 Day of Regression For Empirical Importance

When running a regression of the 5-day realized volatilities on the ES1 5-day realized volatility we get the following results.

```
                                 OLS Regression Results
=======================================================================================
Dep. Variable:      ES1 Index_Ann_Vol   R-squared (uncentered):                   0.841
Model:                            OLS   Adj. R-squared (uncentered):              0.841
Method:                 Least Squares   F-statistic:                              3837.
Date:                Sun, 04 Dec 2022   Prob (F-statistic):                        0.00
Time:                        14:03:00   Log-Likelihood:                          8531.7
No. Observations:                4344   AIC:                                 -1.705e+04
Df Residuals:                    4338   BIC:                                 -1.701e+04
Df Model:                           6
Covariance Type:            nonrobust
==========================================================================================
                         coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------
TU1 Comdty_Ann_Vol     0.4256      0.162      2.629      0.009       0.108       0.743
TY1 Comdty_Ann_Vol    -0.5354      0.055     -9.655      0.000      -0.644      -0.427
TWI USSP Index_Ann_Vol -0.1534     0.039     -3.917      0.000      -0.230      -0.077
LQD US Equity_Ann_Vol  0.5000      0.028     17.979      0.000       0.445       0.554
CL1 Comdty_Ann_Vol     0.0047      0.007      0.715      0.475      -0.008       0.018
Vix Index              0.0038   6.8e-05     55.472      0.000       0.004       0.004
```

Figure 4. Macro variable volatility and VIX as dependant variables for the ES1 volatility.

We can see that all of our variables besides CL1 which is crude oil were significant in helping to improve our estimate of the ES1 realized volatility over the same span. Next, we use the EWMA of these variables in order to see if we can improve upon the above R squared value of .841.

```
                                 OLS Regression Results
=======================================================================================
Dep. Variable:      ES1 Index_Ann_Vol   R-squared (uncentered):                   0.916
Model:                            OLS   Adj. R-squared (uncentered):              0.916
Method:                 Least Squares   F-statistic:                              6786.
Date:                Sun, 04 Dec 2022   Prob (F-statistic):                        0.00
Time:                        14:03:04   Log-Likelihood:                          9920.5
No. Observations:                4344   AIC:                                 -1.983e+04
Df Residuals:                    4337   BIC:                                 -1.978e+04
Df Model:                           7
Covariance Type:            nonrobust
==========================================================================================
                         coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------
ES1 Index_Ann_Vol      8.9862      0.156     57.711      0.000       8.681       9.292
TU1 Comdty_Ann_Vol     1.8139      1.174      1.546      0.122      -0.487       4.115
TY1 Comdty_Ann_Vol     0.4546      0.524      0.868      0.386      -0.572       1.482
TWI USSP Index_Ann_Vol 0.4793      0.325      1.477      0.140      -0.157       1.116
LQD US Equity_Ann_Vol  1.0859      0.217      4.996      0.000       0.660       1.512
CL1 Comdty_Ann_Vol    -0.2254      0.050     -4.546      0.000      -0.323      -0.128
Vix Index             -0.0015      0.000    -11.867      0.000      -0.002      -0.001
```

Figure 5. Macro variable EWMA volatility and VIX as dependant variables for the ES1 volatility.

Our significant variables changed significantly and our overall model R-Squared improved to .916 so this is a more informative set of features than just using the realized volatilities themselves. This makes sense as the EWMA places emphasis on more recent information.

This looks great however, it is not all that useful in terms of forecasting the future realized volatility, for now, we have just seen that there are statistically significant relationships between our independent and dependent variables in real time, but whether we can predict any future values.

2.1 Forecasting Week Ahead Realized Volatility

The final goal of the analysis is to use some variables in order to make a forecast of the week ahead's realized volatility. So for this analysis, we use the same setup, but shift the ES1 realized volatility forward by 5 days in order to find a relationship between the data and the next 5 day's realized volatility. This will be the set up for the XGBoost model as well. This approach in a Linear Regression is a bit naive, but a stepping stone towards using the more complex XGBoost model.

Using the more successful EWMA variables I will attempt to forecast the realized volatility of the next week using the current 10-day EWMA volatility and the implied volatility as given by the VIX.

```
                               OLS Regression Results
=====================================================================================
Dep. Variable:      ES1 Index_Ann_Vol   R-squared (uncentered):                  0.820
Model:                            OLS   Adj. R-squared (uncentered):             0.820
Method:                 Least Squares   F-statistic:                             2816.
Date:                Sun, 04 Dec 2022   Prob (F-statistic):                       0.00
Time:                        14:11:16   Log-Likelihood:                         8257.0
No. Observations:                4339   AIC:                                -1.650e+04
Df Residuals:                    4332   BIC:                                -1.646e+04
Df Model:                           7
Covariance Type:            nonrobust
=======================================================================================
                             coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------
ES1 Index_Ann_Vol          1.2115      0.228      5.313      0.000       0.764       1.659
TU1 Comdty_Ann_Vol         6.8279      1.718      3.975      0.000       3.460      10.195
TY1 Comdty_Ann_Vol        -6.5939      0.767     -8.600      0.000      -8.097      -5.091
TWI USSP Index_Ann_Vol    -2.7592      0.475     -5.804      0.000      -3.691      -1.827
LQD US Equity_Ann_Vol      6.1993      0.318     19.489      0.000       5.576       6.823
CL1 Comdty_Ann_Vol        -0.0767      0.073     -1.058      0.290      -0.219       0.065
Vix Index                  0.0032      0.000     17.582      0.000       0.003       0.004
==============================================================================
```

Figure 6. Macro variable EWMA volatility and VIX as dependant variables for the ES1 volatility.

These results seem encouraging at first, but this is the R-Squared for the period on which the model is trained on. When using these coefficients to find forecasts for the period of 2018-2022, which was held out of the regression, the R-Squared drops to .37 as opposed to .79 which was the out-of-sample R-Squared when estimating the day of realized volatility in the prior models. .37 in itself is not a terrible result, but more importantly, we need to examine the vol spike of 2020 and see how the model performed. The model did forecast rising volatility during this period but was delayed in the level of elevation. It also did not react quickly enough in realized volatility falling back.

Below we look at the performance of the model in and out of the sample, we can see it does pick up on the Covid vol spike, and other times it elevates during times of high volatility but does not capture the magnitude, which can be expected as 5-day volatility is extremely noisy. The forecasts have been moved to the day they are forecasting so the lines are directly comparable.

Below we take a closer look at Covid and see the model was forecasting higher volatility correctly, but not the magnitude and did not capture how quickly the volatility would dissipate.

We do not attempt to use the correlation features for this regression as there is no intuitive reason to suspect a linear relationship or even an interaction relationship between variables and this is not a data-mining exercise.

3. XGBoost Model

As in the prior section we will be forecasting the next 5 days realized volatility. When referring to a date's realized volatility that is the realized annualized volatility for the past 5 days. In all charts and comparisons on a given date we have the realized volatility of the past 5 days and then the prediction by XGBoost which was based on the data 5 days prior in order to predict what the past 5 day realized volatility would be on the given date. So all predictions are truly out of sample predictions and there is no leakage of data which is often the case in many volatility models I have come across.

3.1 XGBoost Methodology

XGBoost is a tree-based ensemble method that relies upon the Gradient Descent algorithm to minimize residual errors as it builds subsequent models. It is more suitable for small and medium-sized data sets than Neural Networks, which would be the go-to choice for supervised learning in large, unstructured data. A boosting algorithm takes the predictors, considered weak learners, and converts them into strong learners. Subsequent trees give extra weight to points incorrectly predicted by earlier predictors. The boosting algorithm will run for M boosting iterations and at termination, it will take a weighted vote from the predictors.

The steps for a boosting iteration are as follows:

1. Let F_m predict a labeled variable $y(F(x) = \hat{y})$.
2. F_m will be associated with a residual $f_m(x) =$

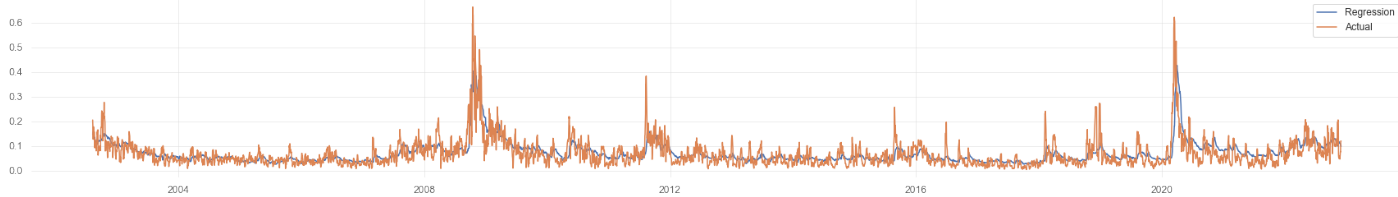

Figure 7. EWMA Model Forecasts vs Actual Realized Volatility

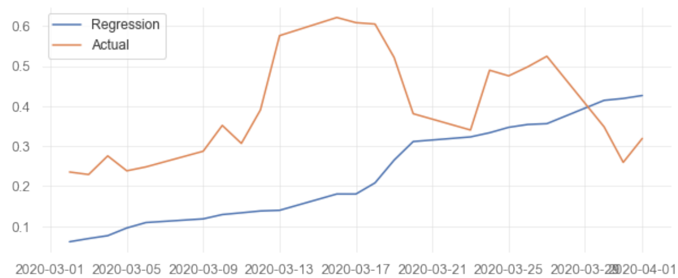

Figure 8. EWMA Model Forecasts vs Actual Realized Volatility during March-May 2020.

$y - F_m$ that fits the residuals in the previous step.

3. Combine F_m and $f_m(x)$ to give $F_{m+1}(F_{m+1} = F_m + f_m(x))$.

F_{m+1} is the boosted version of F_m and will have a lower loss function value than the previous model. New trees are constructed in this fashion for M iterations.

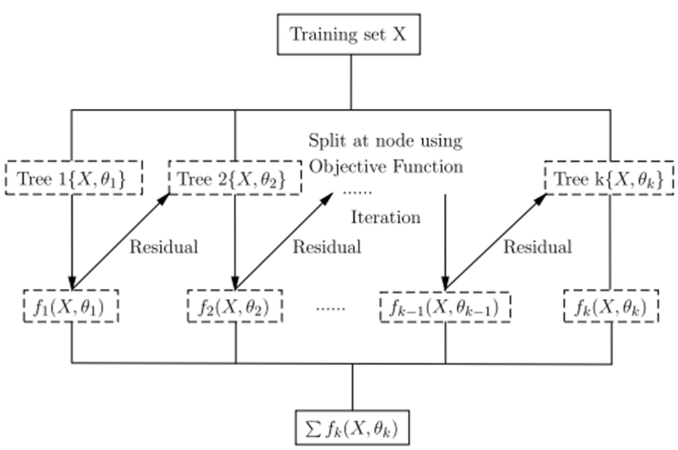

Figure 9. Extreme gradient boosting (XGB)

3.2 Hyperparameter Tuning

XGBoost self-regularizes by introducing L1 and/or L2 penalties and "tree pruning" to prevent overfitting. These two parameters known as alpha and lambda, as well as the number of trees, grown, max depth of trees, and learning rate, were tuned for every rolling prediction in the final model, along with the amount of training data to use to predict each point. This is important because older data may become irrelevant to future data predictions as the distribution changes. The prior three weeks is used for the validation set on a rolling basis for each prediction.

3.2.1 Feature Importance

XGBoost directly evaluates feature importance shown below in we can see the feature importance of the model and how the importance changes throughout time. This is a bit hard to decipher, but clearly there is a wide variety of feature importances throughout time. To give a more intuitive look below are the feature importance for October 4th 2017 vs the feature importance for December 2nd 2022. Feature Importance is given as the number of tree splits on a feature divided by total splits.

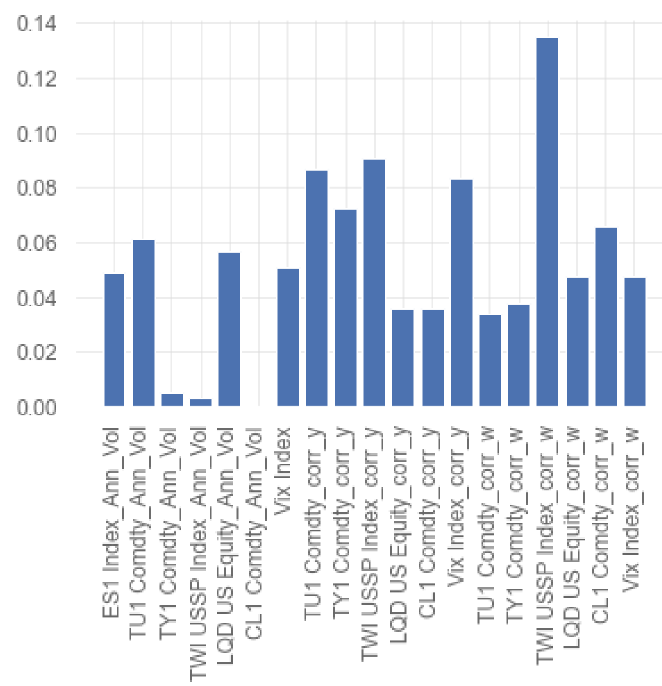

Figure 10. Feature Importance of the XGBoost model in 2017.

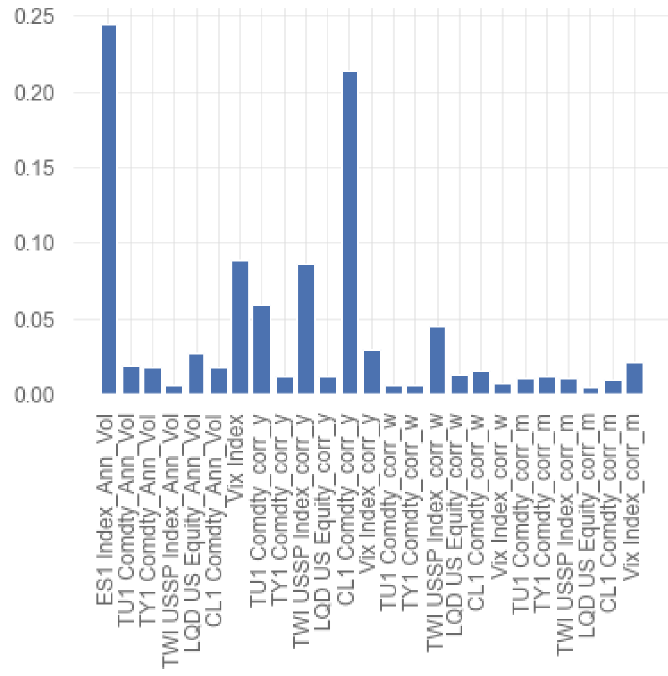

Figure 11. Feature Importance of the XGBoost model in 2022.

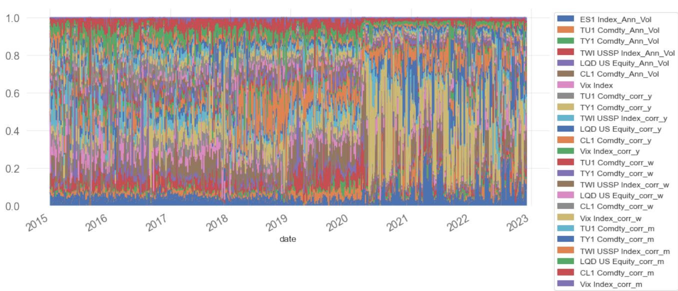

Figure 12. Feature Importance of the XGBoost model throughout the entire testing period.

What we can see from the rolling feature importance is that the model automatically determines which feature is more relevant as time goes on, rather than having many models which must be chosen from depending on a macro regime, XGBoost can build the appropriate model at each point in time and ignore the features which may be less useful, something that linear regression does not do.

3.2.2 Training Length

Another interesting attribute of this model is the ability of the Optuna optimizer to choose whether to use short dated or long dated data for training. I gave the model the ability to choose anywhere from one month to the entirety of data back to 2003 for training a model. Below is how the training data changed throughout time. Some models usef all the data as we can see by the gradually increasing top of the distribution, but also many models were trained on very short or intermediate time spans depending on what was better for correclty predicting the current volatility regime.

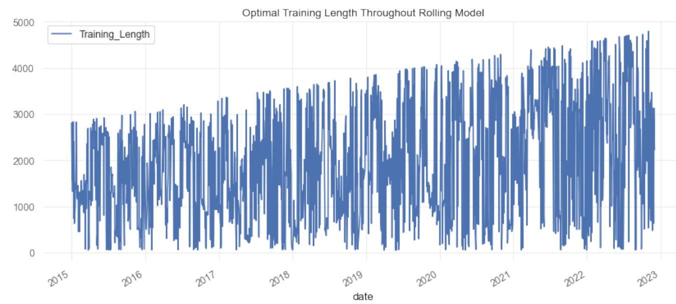

Figure 13. training data length for the model trained to predict a certain date.

4. Model Review

4.1 Covid Vol Spike

The XGBoost model was able to quickly predict the week ahead volatility spike with very little lag. The Covid spike is especially hard to predict given its sudden and external genesis. The model began to show rising volatility before the spike actually occurred but it was less than 5 days before the initial spike which is why a lag can be seen on the chart. It predicted the level remaining high for multiple weeks and was decently accurate on the trajectory of the volatility receding.

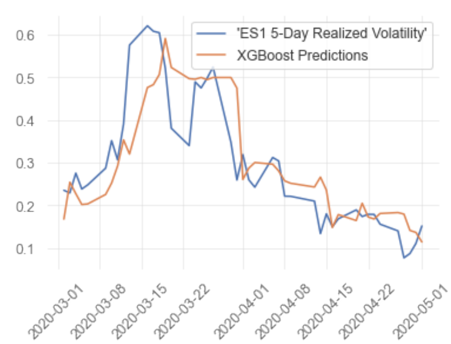

Figure 14. Model performance during the Covid Vol Spike.

4.2 Final Results

When viewing the below graph of actual vs predicted results we can see the amplitude of

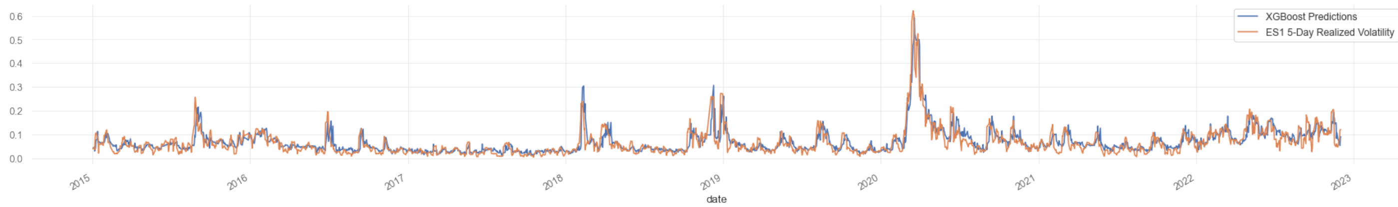

Figure 15. XGBoost rolling model prediction of 5-day realized volatility vs actual.

the spikes was much more accurate. This along with an out-of-sample R-Squared value of .73, the highest among models, proves the viability of this tree-based regression approach. The macro features and the various correlations with the asset volatility target enable some degree of volatility spike forecasting to take place. Looking at the entire period below and the prior close up of the Covid spike, it appears to be a sound model which is not just predicting volatility after the fact, which some of the most 'state of the art' volatility models tend to do. Overall this experiment seems to be a success and has practical application.