Volatility Based Portfolio Allocation

Madison Poore

I. Question & Motivation

Understanding current market volatility (and predicting future market volatility) helps investors make decisions regarding the risk profile of their portfolio. The VIX Index is a 30-day expected volatility of the S&P 500, which is commonly understood as a barometer of investor fear and uncertainty[1]. Therefore, understanding the VIX Index is critical for understanding market dynamics. In practice, a high VIX value indicates a high volatility of the market, meaning the market is fluctuating drastically and (likely) unpredictably, creating unnecessary risk for the investor. To hedge against this risk, many investors may reallocate high risk assets (equities) to lower risk assets (fixed income). If the investor can properly understand and predict market volatility, they will more accurately protect themselves from unnecessary risk. The goals of this paper are as follows: predict the VIX index for 10 observations in November 2025, accurately describe which model best fits historical VIX data, and apply results to a simple theoretical portfolio.

Time Series is a particularly powerful tool here because the stock market has an underlying cyclical nature (due to the market cycle[2]) that is affected by shock events (political, economic, and cultural). The properties of typical market trends provide an opportunity to leverage a Time Series model. Specifically, the auto-regressive variables in the model will contribute to pattern recognition and prediction based on the observed data. The inevitable shock events will be captured by the moving average portion of the Time Series model. Both the auto-regressive terms and moving-average terms combine to create a robust time series model, ARIMA (Auto-Regressive Integrated Moving Average), which will describe stock market behavior.

---

[1] Understanding VIX
[2] Market Cycles

While ARIMA is a strong start to implementing a Time Series model, we will finish by implementing the Prophet Model[3] and a Random Forest Machine Learning Model[4]. Implementing the Prophet Model will adjust for changes in the VIX after trading hours, where the VIX ETF will not describe until the market is open for trading. Finally, a Random Forest will be implemented because Machine Learning tends to capture non-linear behavior of data (in a way that the linear ARIMA model cannot). The presence of these three models and the context of the data motivate the research question:

How has market volatility, as measured by VIX, evolved over the last 10 years? Can an ARIMA model, Prophet Model, or Random Forest Machine Learning Model best describe volatility levels?

II. Statistical Plan

Recall our goals from the **Question & Motivation** section – prediction, pattern recognition, and real-world application. Throughout this paper, each of these goals will be addressed in depth. This section provides a brief overview of the methodology used in this analysis and outlines what follows.

1. Exploratory Data Analysis
    a. To begin, the VIX data will be cleaned and visualized within the Exploratory Data Analysis section. Next, the data will be transformed if necessary, and we will perform structural checks to begin analysis – such as stationarity, homoscedasticity, and correlation between observations.
2. Statistical Analysis
    a. ARIMA
        i. The first part of the Statistical Analysis section will reference the ARIMA model. The model will be optimally fit via the "auto.arima()" function in R, where the final ARIMA model will minimize the AIC statistic. Next, we will evaluate residuals and calculate predictions.

[3] Prophet Model - Meta
[4] Random Forest

The optimal ARIMA model will be explicitly written, with coefficients interpreted and relevant conclusions.

b. Prophet Model

    i. The New York Stock Exchange (NYSE) does not trade securities on weekends and bank holidays, meaning we have holes in our data (depending on how we define a time step in the context of the model). A prophet model is powerful here because it allows us to define our time step as each calendar day and adjust for holidays and weekends where there are no observations of the VIX ETF. While the VIX ETF will not adjust while the market is closed, the (non-tradeable) VIX will adjust since it is a live update of market volatility based on the S&P 500[5]. Finally, the best performing Prophet Model will be described and interpreted within the context of the problem.

c. Random Forest Machine Learning

    i. The last model being considered is the Random Forest Machine Learning Model. Generally, machine learning models perform well when tasked with observing non-linear patterns of data. Due to market cycles, financial data is inherently cyclical (and not necessarily seasonal), making machine learning a critical technology to test and use for statistical inference. Hence, the Random Forest model is the last model, and the most complex, that will be fit within this paper.

3. Final Model & Conclusions

    a. After fitting the three previously mentioned models, we will begin the process of choosing the best performing model. We will leverage the AIC statistic and quantify residuals to determine which model fits historical data the best. Subsequently, we will quantify which model performed best in the lens of

[5] After Hours VIX Adjustments

prediction. This will be done by calculating Mean Squared Prediction Error (MSPE).

4. Applications
    a. Through the lens of the final model, the VIX will be used to aid in portfolio allocation where the goal is to minimize unnecessary risk taken on by the investor.

III. Exploratory Data Analysis

The VIX data used throughout this paper is collected from Yahoo Finance and implemented through the "quantmod" R package[6]. The data set contains adjusted closing price for the VIX ETF (exchange traded fund), which is a tradable fund representing the VIX Index, for all days the market was open in the last 10 years. The data contains 2515 observations. While the quantmod R package provides far more historical data on the VIX ETF, this paper will study the index over the last 10 years because it provides enough time for the VIX ETF patterns to revert to the mean rather than observing a specific economic climate[7]. As discussed in the **Question & Motivation** section, the VIX index represents volatility in the stock market, and is used in practice to speculate on whether volatility will rise or fall[8]. While some trade the VIX directly, most people use information from the VIX to justify other investment decisions[9]. See **Figure 1** for a visualization of the raw data and subsequent discussion.

[6] Yahoo Finance Data
[7] Reversion to the mean in VIX
[8] VIX in practice
[9] VIX ETF

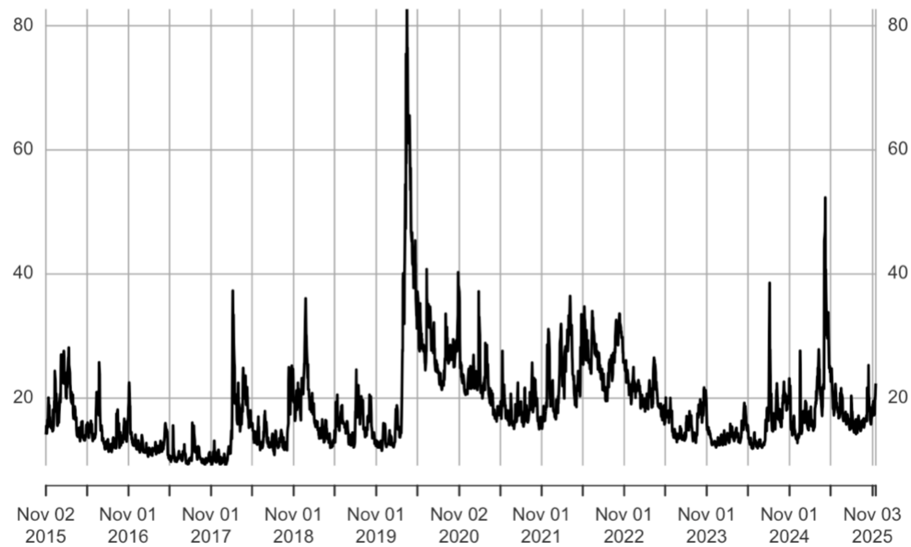

Figure 1: Volatility Index (VIX) 10 year

As seen in Figure 1, the range of observed values of the VIX in the last 10 years is near zero to over 80. VIX values over 30 indicate “greater market fear and uncertainty”, while values below 20 suggest stability of the market[10]. In Figure 1, the reader can see frequent fluctuations between fear and stability, suggesting the data is cyclical and *not* seasonal (since the length of market cycle varies).

One of the most important properties of time series data to leverage is correlation between observations. The ARIMA Model and Prophet Model rely heavily on statistically significant correlation between observations to forecast data, making correlation a necessary condition to test. Visually, we can see significance of correlation between lag terms through ACF/PACF plots. These plots are shown in **Figure 2** and **Figure 3**.

[10] VIX Values Intuition

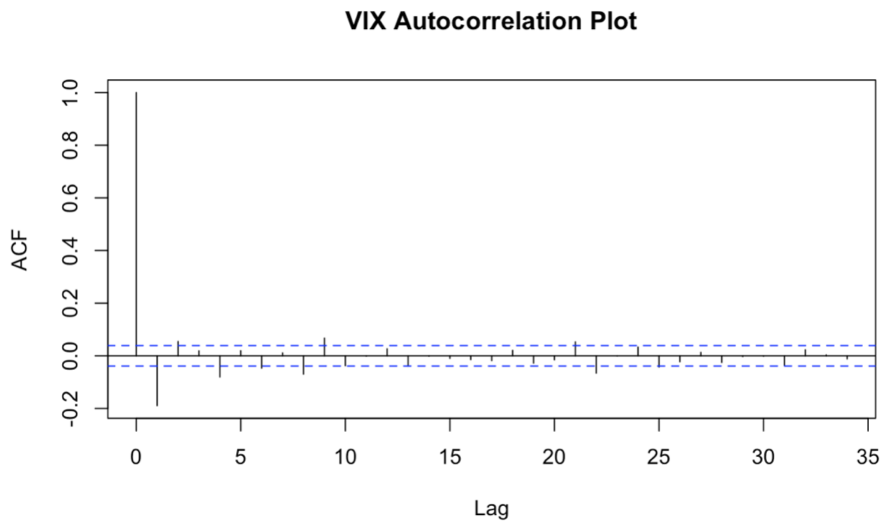

Figure 2: Autocorrelation Plot of Raw VIX Data

Significant terms represented in the ACF plot are the lag terms whose magnitude extends beyond the blue dotted lines, indicating significant autocorrelation of variables. The Autocorrelation (ACF) Plot shown in **Figure 2** indicates we have 3 significant lag variables. That is, the previous three observations significantly impact the current state of the VIX ETF and help motivate the need for a time series model. Within the ACF plot, there is also a wavy pattern, which typically indicates seasonality (or cyclicality). Since these lag variables oscillate with different wavelengths across the ACF plot, there is not seasonality that is easily observed in the data, but rather *cyclicality*. Seasonality is often described as a regular, periodic change in the mean of the series[11]. The data does not show definitively that the VIX will spike at a constant interval, thus we conclude the VIX presents *cyclicality* and not seasonality. Since the data appears cyclical and not seasonal, we will not be able to remove seasonality from the data prior to fitting the ARIMA model and Prophet Model. This gives motivation for the Random Forest Machine Learning model since there is cyclical information that cannot be captured by a linear time series model (like ARIMA), and there may be information in the residuals of linear models that may be quantified to the investor's advantage. The autocorrelation plot is frequently used to provide intuition for

[11] Seasonality definition

determining the p in $ARIMA(p, d, q)$, which is representative of the auto-regressive term (this will be discussed in depth in the **ARIMA** section: **IV. a.**).

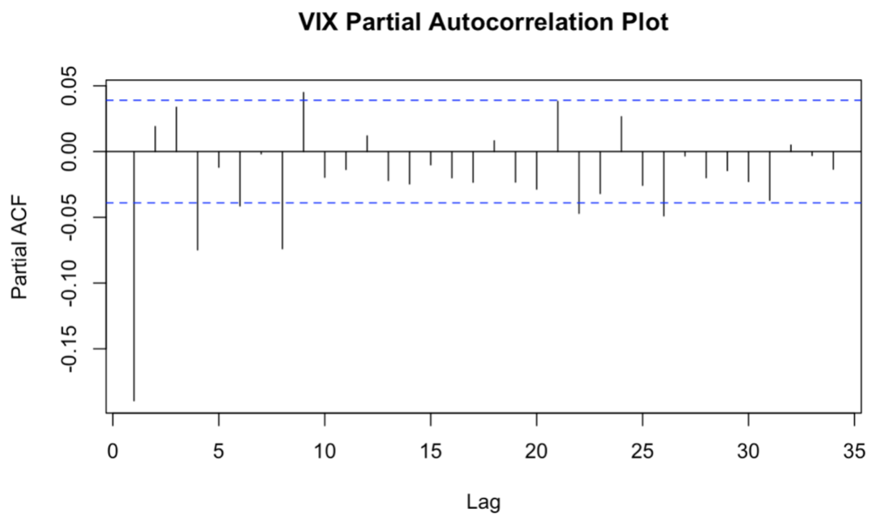

Figure 3: Partial Autocorrelation Plot of Raw VIX Data

Figure 3 represents the Partial Autocorrelation Plot of the raw VIX data, also referred to as the PACF plot. The PACF plot is often used to provide intuition for q, which is used in $ARIMA(p, d, q)$ to represent how many moving-average terms are present. There is one significant lag variable represented in the PACF plot that extends beyond the blue confidence bands. However, we again see the PACF fluctuate, which indicates cyclicality of the data. The cycles of the PACF lag variables are not consistent, however, indicating there is not strict seasonality that may be leveraged in a linear model.

Now that correlation between observations has been established, we must test for stationarity – also referred to as a constant mean throughout the data. Referring to **Figure 1**, there is likely non-stationarity in the raw data because of the drastic spikes and falls of observations. If the data does, in fact, test positive for non-stationarity, an ARIMA model requires differencing of the data before we begin our analysis (to reach stationarity). A statistical test that will determine whether the raw data is stationary or non-stationary is the KPSS test. The hypotheses for the KPSS test are as follows:

$$H_0: trend\ stationarity$$

$$H_a:\ nonstationarity\ (unit\ root)$$

```
{r}
#KPSS test
kpss.test(vix_adj,null=c("Trend"))

 Warning: p-value smaller than printed p-value
        KPSS Test for Trend Stationarity

 data:  vix_adj
 KPSS Trend = 1.5842, Truncation lag parameter = 8, p-value = 0.01
```

Figure 4: KPSS test for Stationarity on Raw VIX data

After running the KPSS test on the raw data, we have a significant p-value (of 0.01, assuming $\alpha = 0.01$), indicating we have statistical evidence to reject the null hypothesis in favor of the alternative hypothesis. In short, the raw VIX data is non-stationary, and we must difference the data before analysis to reach stationarity. After differencing the data, the KPSS test was run one more time to test for stationarity.

```
#KPSS test on differenced time series
kpss.test(diff(vix_adj),null=c("Trend"))

 Warning: p-value greater than printed p-value
        KPSS Test for Trend Stationarity

 data:  diff(vix_adj)
 KPSS Trend = 0.0078771, Truncation lag parameter = 8, p-value = 0.1
```

Figure 5: KPSS Test for Stationarity on Differenced VIX Data

The KPSS test did not return a significant p-value on the differenced VIX data. Therefore, we fail to reject the null hypothesis and can conclude stationarity of our data after it has been differenced once. **Figure** 6 depicts the Differenced VIX index.

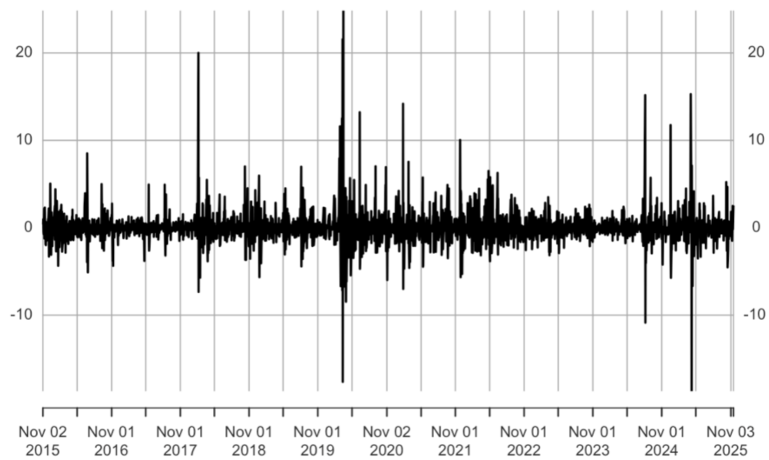

Figure 6: Differenced VIX Index

While the KPSS test explicitly tests for stationarity, but stationarity also requires homoscedasticity over time. Therefore, the conditions of an ARIMA model are met on the differenced data – homoscedasticity, stationarity, and autocorrelation.

IV. Statistical Analysis

The **Data Exploration** section ensured the VIX data met the necessary requirements to fit a Time Series model, like ARIMA. Throughout the next three sections, the ARIMA, Prophet, and Random Forest models previously discussed will be fit and interpreted.

a. ARIMA

The ACF/PACF plots in **Figure 2** and **Figure 3** indicate we likely have a need for both auto-regressive and moving-average terms that the ARIMA model offers. Therefore, the first model discussed in this paper is $ARIMA(p, d, q)$. Conveniently, the "forecast" R package provides a function "auto.arima()" that iterates through potential values of $p, d,$ and q, then compares the AIC statistic between all models tested to find the optimal model (which model minimizes the AIC statistic). This function found an optimal solution of $ARIMA(3,1,1)$. The optimal ARIMA model has three auto-regressive terms ($p = 3$), is differenced once ($d = 1$), and has one moving average term ($q = 1$). **Figure 7** quantifies the magnitudes of each coefficient while providing associated p-values to determine statistical significance of each term.

```
z test of coefficients:

     Estimate Std. Error  z value  Pr(>|z|)
ar1 -1.011301   0.065208 -15.5088 < 2.2e-16 ***
ar2 -0.122664   0.030916  -3.9677 7.257e-05 ***
ar3  0.065160   0.020916   3.1153  0.001837 **
ma1  0.832678   0.062716  13.2770 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 7: Coefficient Significance of ARIMA(3,1,1)

Figure 7 indicates that all three coefficients determined by the optimal ARIMA model are significant. That is, there is statistical evidence that each coefficient impacts the outcome of the model. Therefore, the model may be written as:

$$\Delta VIX_t = -1.01\Delta VIX_{t-1} - 0.123\Delta VIX_{t-2} + 0.065\Delta VIX_{t-3} + 0.83\Delta w_{t-1} + \Delta w_t$$

The Δ comes into play due to the differencing discussed in the model fit. The first three terms with ΔVIX_{t-h} are the three autoregressive terms that have been mentioned throughout this paper. The term with w_{t-1} depicts the moving average term, and w_t is the remaining white noise after the model is fit (more on these residuals later). VIX_t is the current value of the series, ΔVIX_{t-1} has a coefficient of -1.011, which is a relatively large negative number suggesting a strong mean reversion. That is, if the VIX was high yesterday, today it is more likely to move downward (and vice versa). ΔVIX_{t-2} is the VIX term from two days ago, which has a coefficient of -0.123. This likely holds the remaining mean reversion term that was not captured from VIX_{t-1}. The final auto-regressive term, VIX_{t-3}, has a small positive coefficient, which shows a slight tendency for patterns three lags back to continue. The first and only moving average component is a positive term Δw_{t-1} with a coefficient 0.833. A large positive moving average term indicates that the series can adjust strongly to unexpected shocks. Further, if yesterday had a large shock event, today's value will account for that. The final term in this model is w_t, which is the white noise term that is remaining after we have described everything we can with an ARIMA model. That is, this w_t is ideally orthogonal to our explanatory variables.

Having interpreted the optimal ARIMA model, the only remaining task is to evaluate residuals. Using the “check residuals()” function in R, **Figure 8** was achieved.

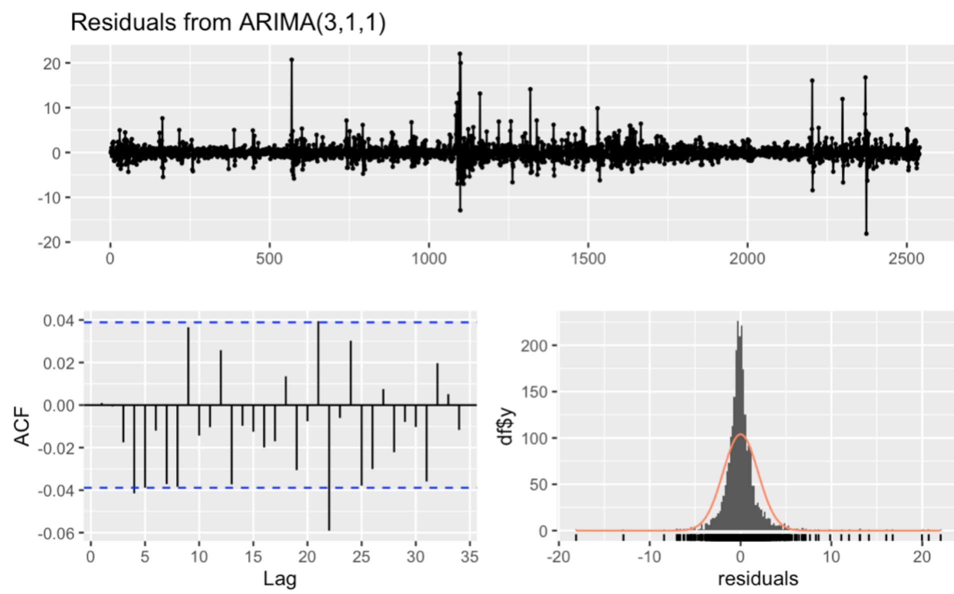

Figure 8: ARIMA Residual Visuals

Within **Figure 8,** the ACF plot indicates there is not significant correlation of residuals, which implies the ARIMA model described correlation between observations properly. The top residual plot mimics the form of the differenced VIX data in **Figure 6**, which solidifies the fact that ARIMA models perform well with linear trends, but are inadequate to analyze this cyclical dataset. One final check to understand the residuals is to visualize a Q-Q plot. This is seen in **Figure 9**.

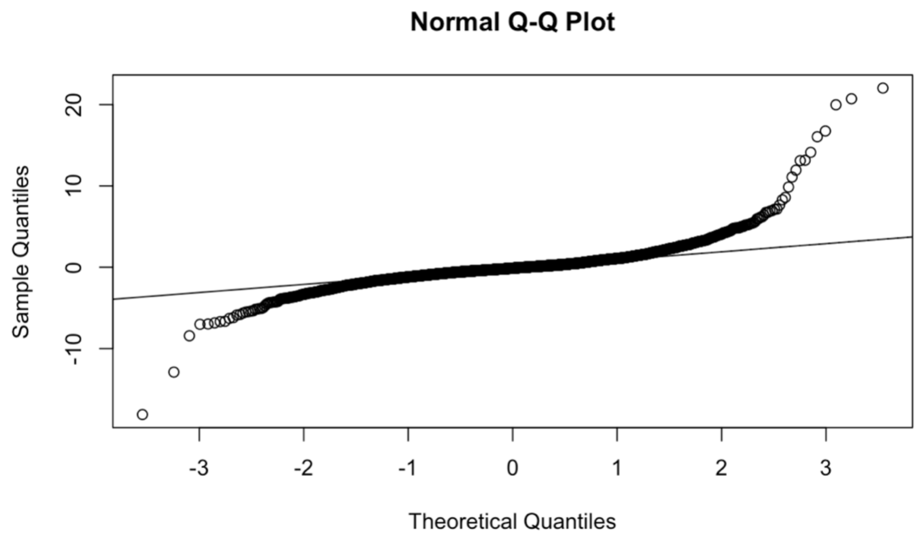

Figure 9: Q-Q Plot of ARIMA(3,1,1) Model

The Q-Q plot in **Figure 9** again proves the abnormality of the residuals from the $ARIMA(3,1,1)$ model. If the residuals were normal, the observed values would appear along the $y = x$ line. Hence, the structure of the tail prompts us to move to a more descriptive model.

Overall, $ARIMA(3,1,1)$ is the best ARIMA model for the VIX data in question. However, the abnormal behavior of residuals and cyclicality of our data lead us towards more complex models (such as the Prophet Model or Random Forest Machine Learning).

b. Prophet Model

While the ARIMA model discussed in **part a** captured short-term autocorrelation in VIX data, it struggled to describe non-linear behavior and shock events. This motivates the use of the Prophet Model[12]. The Prophet model is an additive time series framework developed by Meta that decomposes a time series into four components: a trend component, seasonality, effects of holidays, and residuals. That is,

$$\Delta VIX(t) = g(t) + s(t) + h(t) + \epsilon(t)$$

[12] Prophet Model

Where $g(t)$ represents the trend, $s(t)$ represents seasonality, $h(t)$ incorporates the effects of holidays, and $\epsilon(t)$ is the term representing residuals. The Prophet Model also does not require stationarity of the data, so this model was fit with the initial – raw – VIX ETF data. When fitting in R, it is important to incorporate a calendar of what days the NYSE is not trading to create $h(t)$. The Prophet R package contains a function “holidayNYSE” that includes a calendar for all days the NYSE was not trading in the past, and this is used to help generate $h(t)$. **Figure 10** is a visual of the fit Prophet Model overlayed on the raw data.

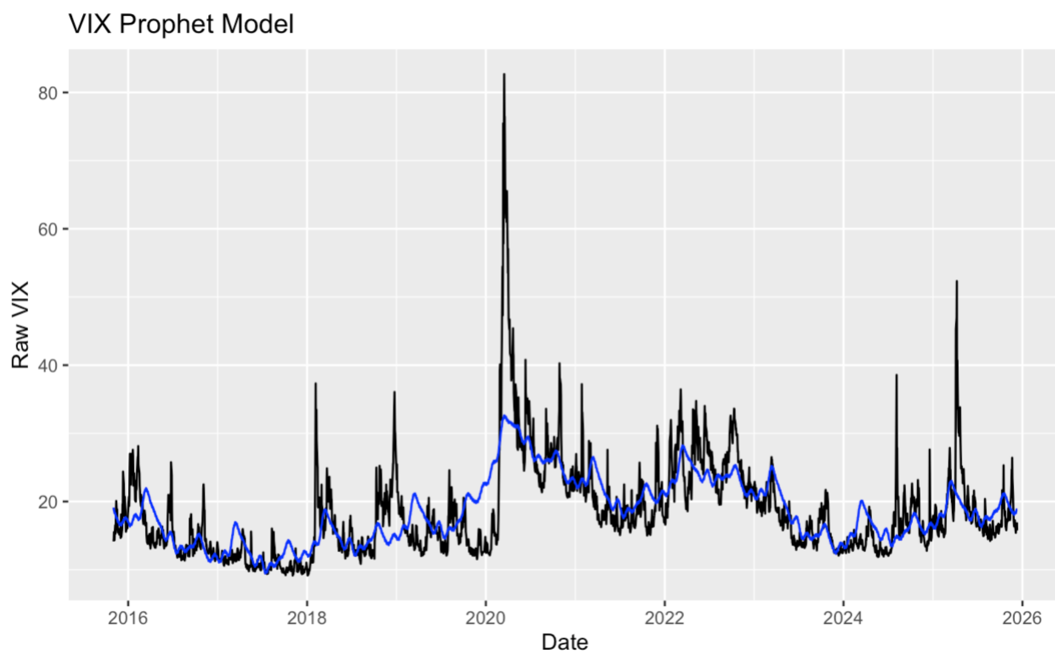

Figure 10: VIX Prophet Model with NYSE Holidays

Visually, the Prophet Model does not accurately capture all broad structural patterns, like the volatility spike in 2020 due to COVID-19. However, the lack of stationarity assumptions in the Prophet Model allows for a more descriptive model than what is expected with an ARIMA model. To understand this model more, it may be decomposed into each of its components as follows.

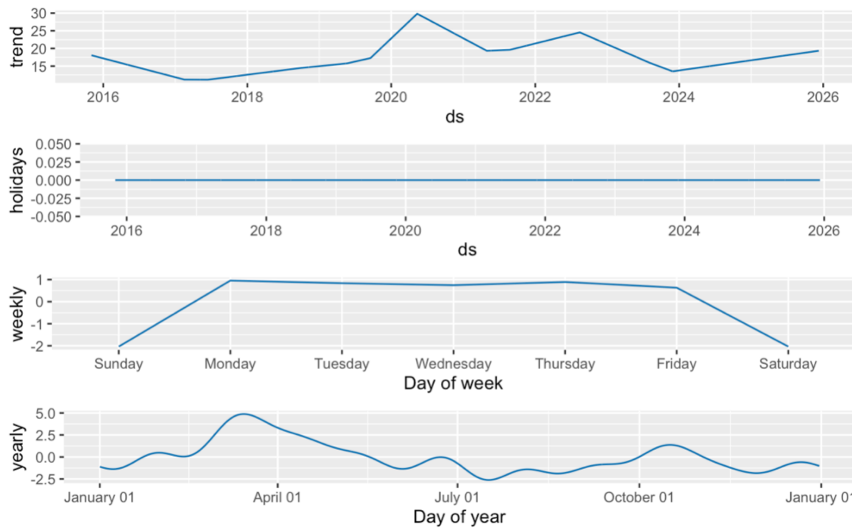

Figure 11: Components of Prophet Model

In **Figure 11**, the fit Prophet Model is decomposed into four categories (representing terms in the general model) – trend, holidays, weekly and yearly seasonality. An important note to make is regarding the holiday's subplot. The flat line in the decomposition of the Prophet model indicates that accounting for holidays does not improve model fit as we would expect. This intuition is reinforced when reviewing the ACF/PACF plots of the residuals (in the following figures).

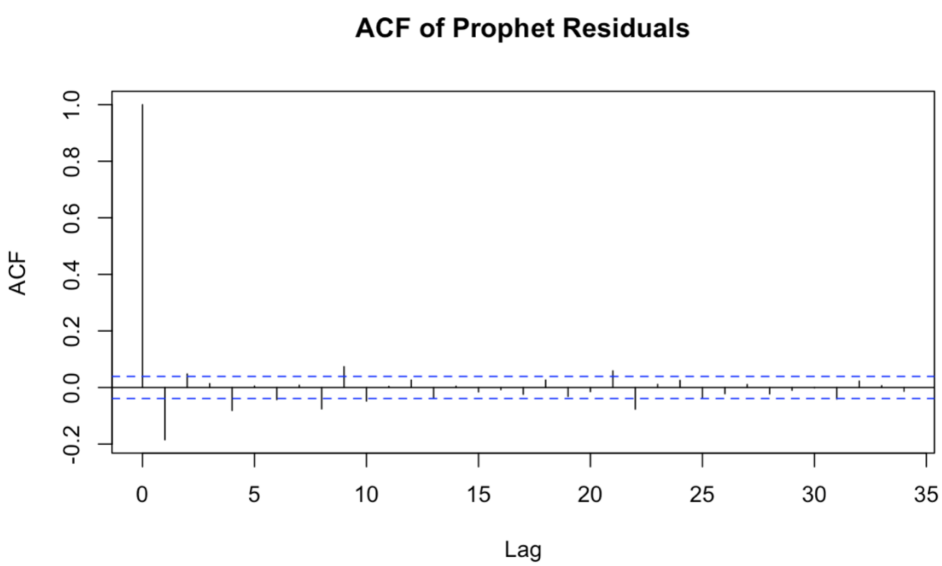

Figure 12: ACF Plot of Prophet Residuals

Like the ACF plot we saw in the ARIMA model, **Figure 12** depicts remaining cyclicality that the Prophet Model could not remove, even with the inclusion of holidays. There are two significant lag variables remaining in the residuals, indicating similar performance to the ARIMA model.

While the complexity of the Prophet model is greater than the ARIMA model, their performance does not substantially differ. While the Prophet Model has the flexibility to describe non-linear behavior within VIX data, the model was unable to capture the cyclicality of the data (even when adjusting for VIX fluctuations after-market hours) and accurately describe the magnitude of historical shock events. Therefore, we are in search of yet another robust model which may more appropriately describe the VIX data. As referenced earlier, the last model fit is a machine learning model, which is known to solve all the concerns with the ARIMA and Prophet models

c. Random Forest Model

The last model that will be fit is a Random Forest Machine Learning Model[13]. A Random Forest is a collection of decision tree models, which averages the predictions of each individual tree to improve accuracy and reduce overfitting. Instead of explicitly modeling time (like ARIMA and the Prophet Model), we create lagged features from past observations to let the model learn short-term dependencies and patterns. This is an explicit benefit of machine learning models within this contes: there is not a need for stationarity, homoskedasticity, correlation of observations, or linearity. Through hyperparameter tuning and feature generation, machine learning models are notoriously good at describing nonlinearity, skewed distributions, outliers, etc.

Technically, there are some key features about Random Forests that must be understood before fitting a model.

Let y_t denote the VIX value at time t, and $X_t = [y_{t-1}, y_{t-2}, \dots, y_{t-lag}]$ be the lagged features (historical observations). Each decision tree in the random forest predicts a value, then

[13] Random Forest

these decision trees are averaged by $\hat{y}_t = \frac{1}{B}\sum_{b=1} \hat{y}_t^{(b)}$, where B is the number of decision trees generated[14]. Each decision tree $\hat{y}_t$ is generated as follows:

1. Row sampling: the algorithm randomly samples N rows (with replacement) from the VIX dataset. Some points may be used multiple times, and some points may not be used at all.
2. Column sampling: At each node in the tree, a random subset of features is chosen, and only those features are considered for tree splitting. Each tree will recursively split by choosing a cutoff value for each feature and running data through it.

The decision tree generation stops when nodes get too small, there's a marginal gain from splitting nodes, or a maximum depth is reached (set by analyst).

Decision Tree algorithms predict by "dropping a lagged feature down the tree", where the feature will land in a leaf. Then we take the mean of the leaf's value. Hence, $\hat{y} =$ (mean of all tree predictions). The averaging of values is what makes the "random forest".

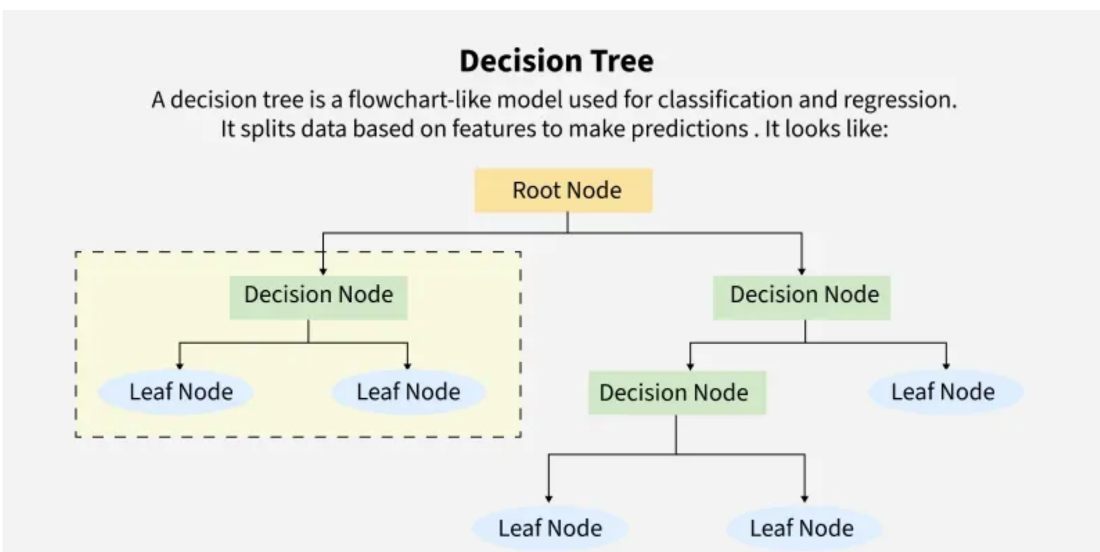

Figure 11: Decision Tree Visualization

As we have discussed in previous sections, a notable strength of Random Forests (and other Machine Learning Models) is the ability to capture non-linear patterns in the data.

[14] Simple Decision Tree Example

Furthermore, machine learning models do not require we have homoskedasticity or stationarity (like we needed in ARIMA). Finally, ML models, like Random Forests can handle missing values well.

Limitations of Random Forests include that it relies on lag features to capture the structure of time, whereas other models have an explicit time component. Second, interpretability is a significant limitation of Machine Learning models. ARIMA and Prophet Models provide a level of explainability that is not easily quantified in machine learning. Finally, machine learning models often require much more data than parametric models to make efficient predictions.

Now that the context of Random Forest generation was provided, the VIX data was fit with this model.

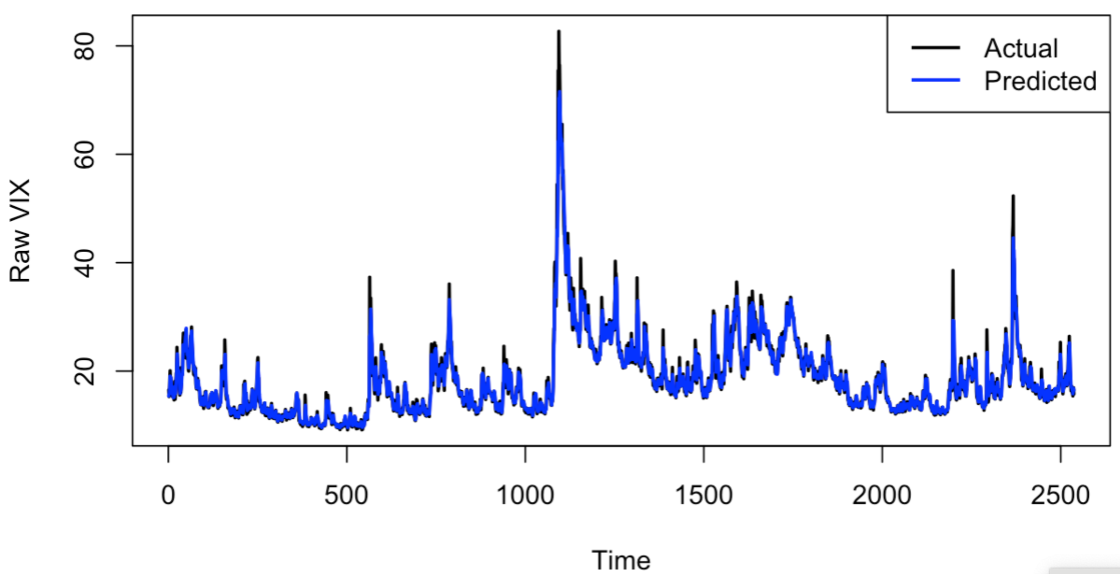

Figure 12: Random Forest Fitted Values vs Predicted Values

Visually, **Figure 12** depicts the most descriptive fit compared to the previous two models discussed. That is, shock events are more adequately described than with the Prophet and ARIMA models, though the Random Forest Model does not entirely describe the magnitude of shock events. Analysis of errors and predictions from the Random Forests Model will be presented in the **Final Model & Conclusions** section.

V. Final Model & Conclusions

To quantify the best performing model, the Root Mean Squared Error (RMSE) will be used. While the AIC statistic provides information for most parametric models with known parameters, it is not adequate to describe a machine learning model such as Random Forests. The RMSE is calculated by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

The best model will minimize the $RMSE$, meaning the fitted values are the closest to the true values[15]. Calculating the $RMSE$ for each fitted model results in the following information:

Model	**RMSE**
$ARIMA(3,1,1)$	1.92598
Prophet	4.99734
Random Forest	1.04304

The model that minimizes the RMSE is the Random Forest Machine learning model, as expected. Therefore, the Random Forest Model is the final (and best) model. As discussed in the preliminaries of this paper, predictions of the VIX perform best on a short-term basis. That is, the further out in time the Random Forest Model is used to predict, the less reliable the predictions become. This is a shortcoming of this analysis – its lack of applicability in long-term portfolio allocations. However, the decision tree model proves reliable in the short-term (often high volume) trading.

VI. Applications

[15] RMSE

Many investors argue for a simple approach to their portfolios, such as the classic 60% of your portfolio in stocks, and 40% of your portfolio in bonds[16]. However, this is a naïve approach once the investor understands the difference of risk profile between stocks and bonds[17]. Stocks are inherently more volatile, while bonds tend to be more stable over the long term. The VIX index provides a measure of stock market volatility, highlighting the risk associated with equities. During periods of high volatility (for example, when the VIX exceeds 30), investors might consider a more conservative portfolio allocation, such as 30% stocks and 70% bonds. If investors could anticipate periods of heightened volatility, they could adjust their portfolios in advance to reduce risk and protect returns.

The previous analysis of the VIX using ARIMA, Prophet, and Random Forest models illustrates that periods of heightened stock market volatility can be anticipated *to some degree*. Among the models tested, the Random Forest Model provided the most accurate in-sample predictions, suggesting that historical volatility patterns contain informative signals. By leveraging these predictions, investors could proactively adjust their portfolios, reducing exposure to stocks when volatility is expected to spike. However, unexpected events, like COVID-19 shock, are difficult to predict, highlighting the limits of *any* model. Future work may explore more advance machine learning techniques or additional market indicators to improve forecasted VIX values and guide a dynamic portfolio strategy.

[16] Basics of Portfolio Allocation
[17] Bonds vs Stocks - Risk Profiling