Predicting Movements of VIX Using Machine Learning and Sampling Methods

ECO 2460 Term Paper

Group Members:

Zhengluo Zhang 1008516688

Min Ho Kim 1003056266

University of Toronto

April 9, 2023

Introduction

The VIX is an index introduced by the Chicago Board Options Exchange 1993 that measures expected short-term market volatility and provides a benchmark for historical levels of market anxiety. Market anxiety refers to a state of unease, nervousness, or worry among investors about the future direction or performance of the financial market. It is characterized by a high level of uncertainty and fear, often driven by economic, political, or social factors that can have a significant impact on the market's stability and investor confidence. Market anxiety can result in increased market volatility, as investors react to new information or changes in market conditions. Thus, the VIX is considered by investors as a measure of investors' risk-aversion and market sentiment and is widely quoted as the "investor fear index". Analogous to the yield curve, the level of the VIX represents the expected future stock market volatility, which is derived from the current prices of options on the S&P 500 Index. The VIX also serves as an index upon which futures and options contracts on volatility can be written. This allows investors to trade these derivatives as a hedge against potential market downfalls. Therefore, the ability to predict the evolutions of VIX is crucial for at least two purposes:

1. Predict movements in investor expectations on market volatility, and

2. Predict movements in prices of tradable VIX derivatives.

The main emphasis of this paper will be on the first point. We utilize machine learning techniques to predict the up-and-down movements of the VIX index. Borrowing from the ideas of Prasad, Bakshi and Seetharaman (2022), we select features that are most important in predicting the movements of the VIX and optimize the model. In addition, we refer to Clark's influential paper (1973) on the subordinated distribution of asset returns and integrate the idea into a time-series model using rejection sampling algorithm. We compare and analyze the relative performance of the models. The performance is measured by the models' respective precision and recall rates. One argument is to show that with the correct augmentations, traditional time-series forecasting models are not necessarily outmatched by more recent and complex machine learning models in predicting financial time-series.

More about the VIX

Value of an option contract depends on many things, including volatility (σ). Implied volatility is the value of volatility that can be derived given an option-pricing models. Mathematically, if $V = f(\sigma, \boldsymbol{\theta})$ is the value of an option contract and f is the option pricing methodology, then assuming there exists an inverse of f such that $f^{-1}(.) = g(.)$ then implied volatility is $\hat{\sigma} = g(V, \boldsymbol{\theta})$; and thus $\hat{\sigma}$ is the volatility that is implied by the value of the underlying option contract. Essentially, implied volatility is the volatility that equates the value of an option pricing model to the option market quote.

Chicago Board Options Exchange (CBOE) is the world's largest options exchange based in Chicago, and facilitates the trade of options, futures, and other financial derivatives. VIX is the volatility index produced by the CBOE since 1993 and it aims to measure the 30-day expected volatility of the stock market in the United States. In doing so, it uses the prices of S&P500 index call and put options to derive the implied volatility described above. VXO is the old VIX, where it benchmarks the S&P100 index (OEX), and has been replaced by the VIX since 2004. Also, the VIX is calculated using market prices whereas VXO is based on the averages implied volatility derived from Black-Schole priced options. It is important to note that the direction of the VIX does not necessarily reflect the direction of the market movement.

There seem to be a couple economical reasons for the switch in methodology. According to Carr and Wu (2006), one of the primary reasons lie within the interpretation of the two measures. The current VIX provides an economical interpretation as the "...price of linear portfolio of options". As we have described mathematically above, the implied volatility computed to construct VXO is just some transformation of option prices and fails to provide any further economic interpretation. Furthermore, they claim that the VXO methodology creates "...an artificial upward bias". This can be inferred from our discussion of the difference in the magnitude of the two series; the variance of VXO is somewhat larger than that of

VIX (98.1 vs 80.5). Also, as mentioned above, the VIX incorporates more firms (S&P 500) and thus captures a better picture of the overall market volatility, compared to the VXO methodology which only uses top 100 firms in S&P 500 in its calculation. That said, VIX is still just an estimate for market volatility, as is the VXO. VXO may still provide useful information for investors in regards to the volatility of the market, and investors will utilize any relevant information to profit from. The restriction to 100 top firms could still be useful when investment decisions solely revolve around top large-cap firms in the S&P100 index.

Figure 1 below illustrates the evolution of the VIX since the start of the measurement. The huge spike between Q1 and Q2 of 2020 reflects the market volatility during the first few months of the global spread of COVID-19. Ex post, we have observed significant asset inflation following Q2 of 2020, which carried on until near the end of 2021 when the Federal Reserve were beginning to hike rates, which was followed by market corrections. Thus, the VIX seem to abruptly peak when there is an unexpected, sudden change in the market condition, which in this case was the COVID-19 pandemic. The volatility to this day remains high compared to levels before the pandemic, and this can be seen as a result of the market correction.

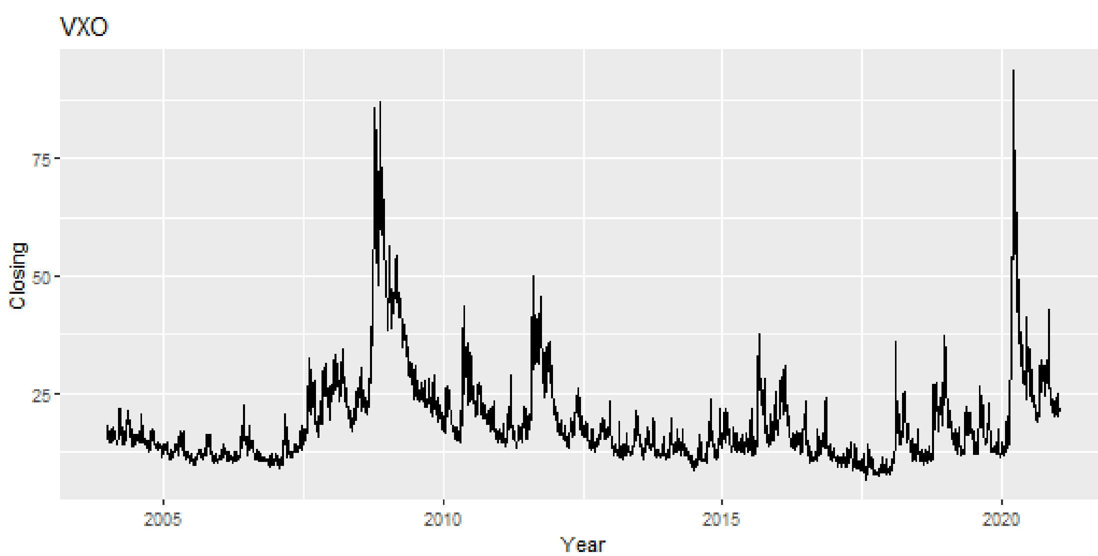

Figure 1: Evolution of VIX 2004 to 2021

Literature Reviews

There are many pre-existing works on predicting the movements of the VIX. More generally, there is a plethora of financial literature that claim that the prediction of short-term volatility is possible. In 1982, Engle proposed the Autoregressive Conditional Heteroskedastic (ARCH) model to estimate the forecast variance, and its impact has been enormous for both practitioners and scholars. Degiannakis (2008) used a variation of the more traditional Autoregressive Moving Average (ARMA) model to predict the VIX.

More recently, advancements in machine learning has led scholars to utilize complex statistical algorithms for time-series predictions. Some of the popular algorithms that have been used in the past include K-Nearest Neighbors (Lahmiri, Bekiros and Bezzina, 2019), Support Vector Machines (Rosillio, Giner and Fuente, 2013), and Multilayer Perceptrons (Bai and Cai, 2023). Some of these methods are used not only for prediction purposes, but also for feature selections. Feature selection is an essential step in the machine learning pipeline, which aims to identify and select the most relevant features from the available set of features to improve the accuracy, reduce overfitting, and increase the speed of the learning algorithm. When developing a machine learning model, it is important to choose the most relevant features, which are the variables that have the most impact on the outcome you are trying to predict. By selecting the most relevant features, you can reduce the complexity of the model, improve its performance, and reduce the risk of overfitting. Overfitting occurs when a model is too complex and starts to fit the noise in the data, rather than the underlying patterns. This can lead to poor performance on new data that the model hasn't seen before. By selecting the most relevant features, you can reduce the risk of overfitting and improve the generalizability of your model. In addition, feature selection can also help to reduce the computational resources required to train a machine learning model. By reducing the number of features, you can reduce the computational burden of the model, making it more efficient and easier to scale.

In regards to this, Prasad et al (2022) have analyzed which macroeconomic variables contribute the most in predicting the movements of the VIX. They employ three different algorithms for this purpose: Logistic Regression, Light Gradient-Boosted Machine (Light GBM), and Extreme Gradient Boosting (XG Boost). All of these models forecast whether the VIX will rise or fall the next time step, encoded as 1 and 0 respectively. They rank the impact variables have in prediction by the feature scores, and they show that the Top 5 most important factors are $VIX_{t-1}, VIX_{t-2}, VIX_{t-3}, VIX_{t-4}$ and P_Gold_{t-3} where P_Gold is the price of gold recorded by the London Bullion Market Association (LBMA). The inclusion of gold price is perhaps not surprising. Gold has long been considered *de facto* hedge against risk and economic downturns. Although this belief has been challenged for some time, investors may still opt for gold in times of large volatility, as indicated by the VIX. We will utilize these five features to perform our own predictions.

Data Processing and Model

Data

The data we use is directly available from the CBOE's website and gold price data is from LBMA. Observations range from the start of 2004 all the way up to start of 2021. We take the difference of each series as we are concerned with the returns of these assets. We encode the returns as either an upward movement (1) or a downward movement (0). Then we split the sample evenly into training and test set. The hyperparameter tuning procedure involves further splitting of the training set into a sub-training set and validation set. By performing a grid search on the possible range of the hyperparameters, we obtain the optimal set of hyperparameter values that we use to evaluate model performance on the test set.

Models

We make use of two of the machine learning algorithms specified above in Prasad's work, with an additional model of LSTM Neural Network as an original extension. The hyperparameters of choice used to optimize each model are listed below:

1. XG Boost: *n_estimators*, *max_depth*, *learning_rate*

2. Light GBM: *n_estimators*, *max_depth*, *learning_rate*, *reg_lambda*

3. LSTM: *neurons*, *dropout_rate*, *optimizer*

A recurrent neural network (RNN) is a type of artificial neural network that is designed to process sequential data by maintaining a hidden state or memory of the previous inputs. LSTM (Long Short-Term Memory) is a type of recurrent neural network (RNN) that is capable of processing sequential data with long-term dependencies. It was introduced by Hochreiter and Schmidhuber (1997) and has since become a popular choice for many applications such as speech recognition, natural language processing, and time series analysis. The main advantage of LSTM over traditional RNNs is its ability to selectively remember or forget information from the past. This is achieved through the use of gates, which are special types of neurons that control the flow of information in and out of the LSTM cell.

The LSTM cell has three main gates. The forget gate controls how much of the previous cell state should be forgotten. The input gate controls how much new information should be added to the cell state. Lastly, the output gate controls how much of the current cell state should be outputted. In addition to these gates, the LSTM cell also has a memory cell that stores the current state of the LSTM. During each time step, the input to the LSTM cell is a combination of the current input and the output of the previous LSTM cell. The input is then passed through the input gate to update the memory cell. The forget gate then determines which information from the previous cell state should be discarded. Finally, the output gate decides which information from the current cell state should be outputted.

We now introduce a fourth model that is unique to our work. To begin, let y_t be defined as the logarithm of VIX level at time t. The sample autocorrelation (ACF) plot in Figure 2 illustrates the long term dependence structure that is present in the data. Traditional time-series modelling requires the stationarity assumption, which in this case does not seem to hold. If the series was indeed stationary, then what we should observe is an abrupt fall its sample autocorrelation. One way to create to transform this into a stationary series is to difference the series. Thus our analysis and outcome variable is in terms of the return of log prices. We can see from the ACF plot that by differencing the series, we have created a new series that is stationary. The autocorrelation drops sharply after h=0 (which is correlation with itself) and almost immediately falls within the confidence interval at around h=2. We observe similar patterns on PACF, which suggests that AR(2) will be the ideal choice to model the sequence.

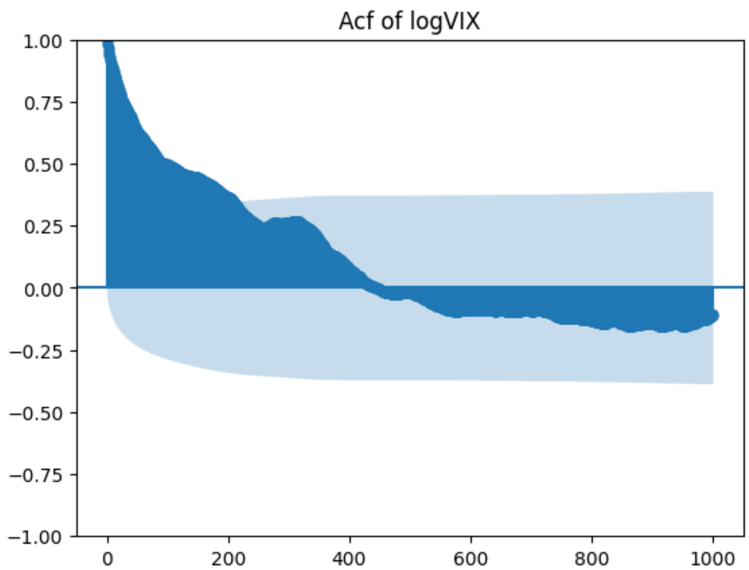

Figure 2: ACF of y_t

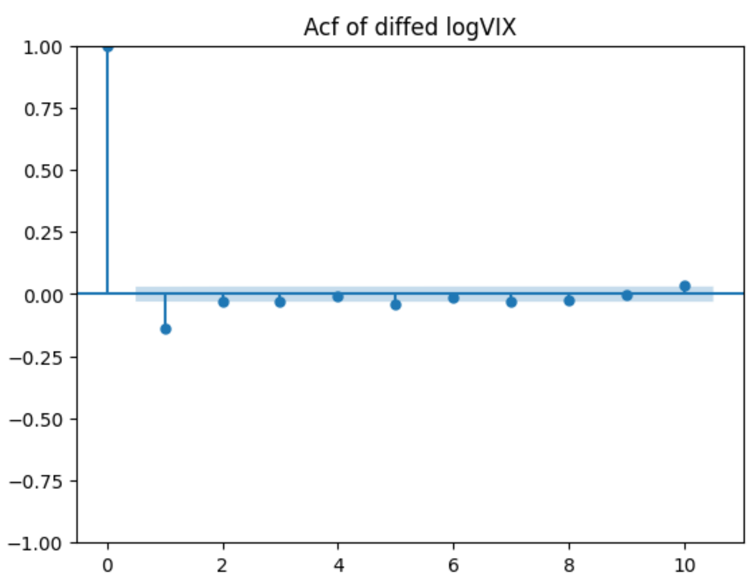

Figure 3: ACF of Δy_t

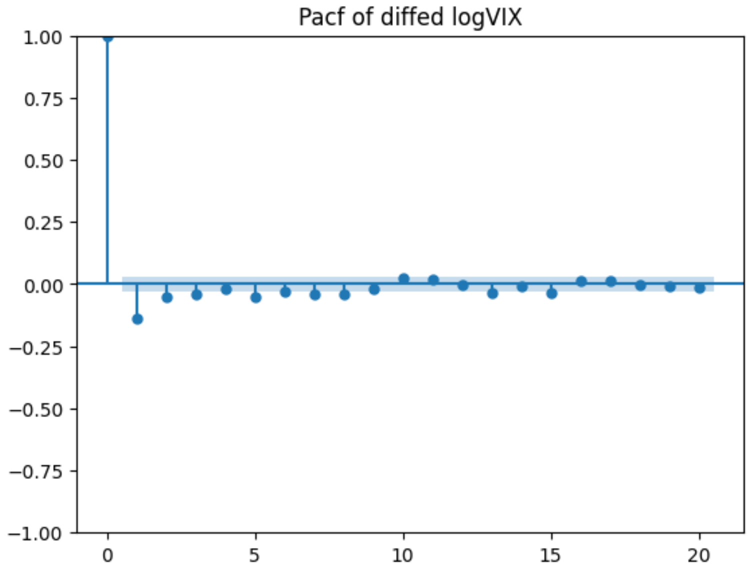

Figure 4: PACF of Δy_t

Now we specify a regression model of the form $\Delta y_t = \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-1}^2 + \epsilon_t$ and run a linear regression. While the coefficients are significant for both lagged terms, the overall prediction performance is very poor ($R^2 \approx 0.02$). The efficient market hypothesis says that the current price contains all the useful information, and best prediction of future price is current price. Specifically, the efficient market hypothesis says that the price of assets should follow a random walk, and thus price change cannot be predicted by past price changes. This is supported by the low R^2 of the regression, despite the significant coefficient estimates, which indicates most of the information in predicting change in asset price still comes from the noise.

One assumption behind this autoregressive model is that it assumes the error terms are normally distributed with mean zero. Traditionally, it was assumed that price increments satisfied the properties $E(\Delta X_t) = 0$ and $E(\Delta X_t \Delta X_s) = 0,\ \ t \neq s$ Thus it was previously conventional to infer a random walk process on the price increments:

$$X_t = X_{t-1} + \epsilon_t \implies \Delta X_t = X_t - X_{t-1} = \epsilon_t$$

where $\epsilon \sim i.i.d\ \ N(0,\sigma)$ is a random walk process.

However Clark (1973) argues that while for some asset price X_t, ΔX_t are independent and they cannot be assumed to be normally distributed, which he proves by showing that the distribution of ΔX_t has a higher sample kurtosis than 3; subsequently he describes the distribution of ΔX_t as being leptokurtic. Subsequently, we cannot apply central limit theorem to ΔX_t. New market information will impact the speed of price change, which translates to higher frequency of trades, and thus prices. Usually X_t would be indexed by discrete time $t = 1, 2, 3, ...$ Clark argues that if t itself is stochastic process with positive increment $T(t)$ such that $t_1 \leq t_2 \leq t_3...$, then we can create a new stochastic process $X(T(t))$, whose distribution is subordinate to the distribution of $X(t)$. $T(t)$ essentially measures the speed of evolution of $X(T(t))$. The following is a summary of the mathematical definition he provides in his paper.

Let $X(t)$ and $T(t)$ be processes with stationary independent increments defined as

1. For $k \in \{1, 2, 3, ..., n-1\}, X(t_{k+1}) - X(t_k)$ (and for $T(t_{k+1}) - T(t_k)$) are mutually independent for any finite set $t_1 \leq t_2 \leq t_3... \leq t_n$

2. $X(s+t) - X(s)$ (and for $T(s+t)) - T(s+t))$ only depends on t, not for any s.

Suppose $\Delta X(t)$ follows some distribution with mean 0 and finite variance $\sigma^2 < \infty$. Also suppose $\Delta T(t)$ follows some positive distribution with mean α and is independent of $\Delta X(t)$. Then $X(T(t))$ is a subordinated stochastic process of $X(t)$, and $\Delta X(T(t))$ is stationary with mean 0 and variance $\alpha\sigma^2$. Furthermore, he shows that if we assume $X(t)$ is normal, and if we condition on $\Delta T(t) = v$, we have that

$$\begin{aligned} &E(\Delta X(T(t))^4 | \Delta T(t) = v) = 3v^2\sigma^4 \\ &\qquad \implies E_{\Delta T(t)}(3v^2\sigma^4) = 3\sigma^4(\alpha^2 + var(v)) \\ &\qquad\qquad \implies k_{\Delta X(T(t))} = \frac{3\sigma^4(\alpha^2 + var(v))}{\alpha^2\sigma^4} = 3(\frac{\alpha^2 + var(v)}{\alpha^2}) > 3\ \ \forall \alpha \in R \end{aligned}$$

Which implies that the distribution of price increments (indexed by the stochastic process

$T(t)$) is indeed leptokurtic. Figure 5 illustrates the sample histogram where we observe a leptokurtic distribution of residuals centred around 0. The residuals clearly are not normally distributed, and we can confirm this from its sample kurtosis of 5.2, which is much bigger than 3 of a normal distribution. Since we know the distribution is following some distribution subordinate to Gaussian, we will apply rejection sampling skill to approach the distribution of the error term, then introduce the simulated noise into our model. The augmented model is :

$$\Delta y_t = \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-1}^2 + \phi_3 \xi_t + \epsilon_t$$

where ξ_t is the simulated noise. To implement the rejection sampling, we choose Laplace distribution($\frac{1}{2b} exp(-\frac{x-\mu}{b})$), since it has exponential tails that decay more slowly than the normal distribution.

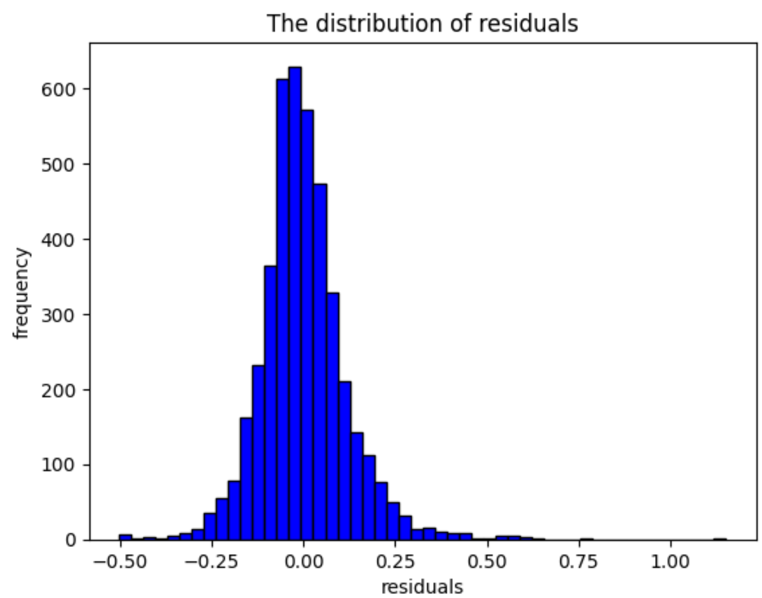

Figure 5: Distribution of Residuals

Rejection sampling is conducted the following way:

1. We generate a sample from the proposed Gaussian distribution $q(x)$.

2. The second, u is generated uniformly from the interval $[0, cq(x)]$

3. If $u >$ the density of Laplace distribution, then x is rejected. Otherwise we accept it.

By doing this, we are not sampling directly from Laplace distribution, but keep the properties of heavy tails compared to normal distribution. For simplicity, here we pick c=5, which is shown in Figure 7. To summarize, we first simulate noise for training and testing set then check the model accuracy by converting the result into 1 and 0. We repeat this step 100 times and finally calculate the average performance.

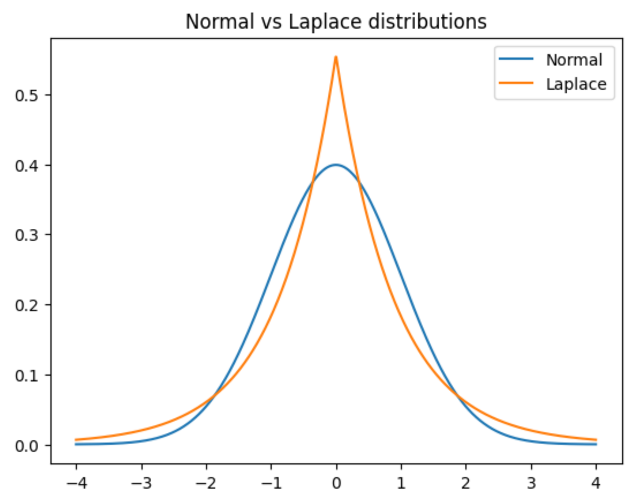

Figure 6: Laplace Distribution Subordinate to Normal Distribution

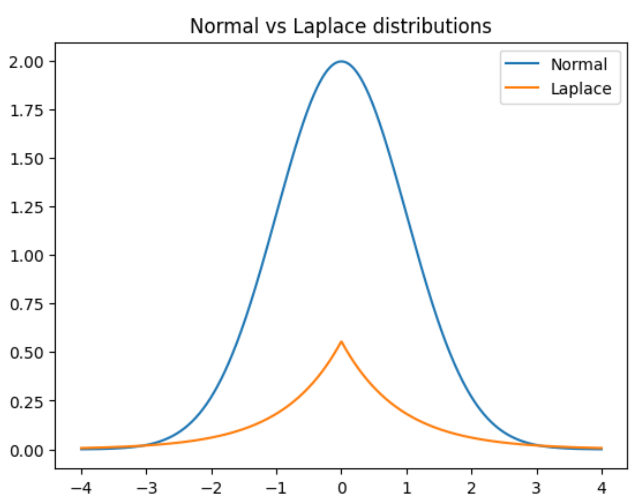

Figure 7: Rejection Sampling $c = 5$

Results

Our main result is presented in Table 1. Accuracy is the number of correct classification made by a model. We can see that there are essentially no differences in terms of the accuracy across all four models. Where we do observe noticeable differences is in the recall rates. Recall is the fraction of correct positive predictions among all actual positive instances. Thus recall measures how many of the positive instances were correctly predicted by the model. In the context of our study, Recall metric of class 0 would be the number of correctly identified downward movements divided by the sum of false positives (predicted up, actual was down) and true negatives (predicted down, actual was down). XG Boost and Light GBM dominates LSTM and Non-linear AR under this metric by almost two-fold margin, but recall rate for class 1 ($\frac{\#TP}{\#TP+\#FN}$) is conversely much higher for LSTM and Non-linear AR models. In terms of precision for class 0, the Non-linear AR model is shown to outperform other models by a considerable amount, while XG Boost show the highest precision for class 1.

Model	Accuracy	Precision (0)	Recall (0)	Precision (1)	Recall (1)
XG Boost	0.53	0.523	0.882	0.571	0.163
Light GBM	0.53	0.524	0.843	0.556	0.204
LSTM	0.52	0.541	0.392	0.508	0.653
Non-linear AR	0.532	0.559	0.395	0.517	0.676

Table 1: Model Performance

Conclusions

In this paper, we applied several machine learning methods and a non-linear AR model to the movements of VIX and compared their results. Though the accuracy is just slightly better than a random model, there are some insights that can be given:

1. *Macroeconomic factors matter even though with low weights in the feature map.* Since introducing more factors will not necessarily increase the performance of the model, we checked the result of only selecting several of the most important variables from the current literature. Since the result is just slightly better than a randomized model, we confirm that those macroeconomic factors are still necessary.

2. *There is some extra information from the error term that we can capture to improve our model.* In our non-linear AR model, we achieve the same accuracy with only factors of past observations. Moreover, the precision-0 and recall-1 are even higher than the other models. This implies that our sampling method does provide some useful information from the true distribution of the noise. Future work can be done by finding more properties of the noise distribution and sampling techniques.

References

Y. Bai and C. X. Cai. Predicting vix with adaptive machine learning. *SSRN Electronic Journal*, 2023. URL `https://ssrn.com/abstract=3866415`.

G. Bekaert and M. Hoerova. The vix, the variance premium and stock market volatility. *Journal of Econometrics*, 183(2):181–192, 2014. ISSN 0304-4076. doi: 10.1016/j.jeconom.2014.05.008.

P. K. Clark. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41(1):135–155, 1973. URL `https://www.jstor.org/stable/1913889`.

S. A. Degiannakis. Forecasting vix. *Journal of Money, Investment and Banking*, (4):135–145, 2008. URL `https://ssrn.com/abstract=1806044`.

R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

S. Lahmiri, S. Bekiros, and F. Bezzina. Multi-fluctuation nonlinear patterns of european financial markets based on adaptive filtering with application to family business, green, islamic, common stocks, and comparison with bitcoin, nasdaq, and vix. *Physica A: Statistical Mechanics and its Applications*, 538:122858, 2020. ISSN 0378-4371. doi: 10.1016/j.physa.2019.122858.

A. Prasad, P. Bakhshi, and A. Seetharaman. The impact of the us macroeconomic variables on the cboe vix index. *Journal of Risk and Financial Management*, 15(3):126, 2022. doi: 10.3390/jrfm15030126.

R. Rosillo, J. Giner, and D. de la Fuente. The effectiveness of the combined use of vix and support vector machines on the prediction of s&p 500. *Neural Computing and Applications*, 25(2):321–332, 2014. doi: 10.1007/s00521-013-1487-7.