A Comparison of Non-linear Forecast Models for the VIX and its Volatility

The volatility parameter is the single most important parameter for pricing short-term options and arguably one of the more important parameters for virtually all financial time series. The market focuses on the VIX or Fear Index, an option-implied forecast of 30 calendar-day realized volatility of S&P 500 returns derived from a cross-section of vanilla options. We examine models that only utilize past values of the VIX and document improvements in forecasting the VIX (and its volatility) over different horizons. The approaches include LSTM models, simple moving average methods, data-driven volatility techniques, and industry models like Prophet, as such methods avoid model identification and estimation issues, especially for a series like the VIX which is non-stationary. We compare models and document which ones perform best for varied horizons.

Additional Key Words and Phrases: Non-stationarity, neural networks, LSTM, Prophet, VIX, volatility, forecasts

ACM Reference Format:

1 INTRODUCTION

In this paper, we examine various approaches to forecasting the VIX, possible the most widely used volatility barometer for stock market risk. It is also known as the “Fear Index.” Risk estimation is a widespread task in the financial services industry [10] and begins with the estimation of market volatility. The VIX is determined using a formula that derives the market’s expectation of realized one-month standard deviation of returns backed out from the near-term call and put options on the S&P 500 index [2]. Far from being just an indicator of index volatility, VIX forecasts are often used to calibrate individual stock volatility forecasts using the Capital Asset Pricing Model (CAPM), so the models in this paper apply to forecasting index volatility and that of individual stocks. These in turn, help drive various risk metrics like value-at-risk (VaR), conditional value-at-risk (CVaR), and expected shortfall (ES).

There are several interesting characteristics of the VIX that make this forecasting problem of VIX and the volatility of VIX (the so called VVIX) an interesting one.

1) The VIX series is non-stationary. Linear time series models for the mean that require differencing for stationarity are poor templates. First, traditional time series models such as SARIMA impose stationarity and normality. Second, some, but not all models assume normality. The work in [5, 11] shows that LSTMs do better in these settings.

2) The sample autocorrelations of the VIX series has hyperbolic decay. For this reason we include an LSTM model and also use Facebook’s Prophet [15] model,1 both of which have design features that will capture non-linearity and non-stationarity by design.

1https://facebook.github.io/prophet/

Author’s address:
(3) The family of GARCH models for conditional variance are a poor choice for volatility forecasting of VIX. Volatility forecasting using traditional time-series methods (such as GARCH, EGARCH, APARCH, GJR and IGARCH, [9]) imposes two constraints, stationarity and distributional assumptions (normal or Student t). However, the nonlinear data-driven neural net (NN) models for volatility (we call these neuro-volatility models) do not require any constraints.

(4) The fact that asset returns are non-normal has important consequences in finance, where assuming normality leads to underestimating risks, often with dire consequences. [1, 6] show that assuming non-normality in VIX forecasting models improves on the normal models used in [4].

(5) Commonly used forecast models are historical simulation (HS), moving average (MA), Normal GARCH, Student-t GARCH, and EWMA for squared continuously compounded returns [1, 3, 4, 7, 9]. Forecasts are obtained of the conditional variance ($\sigma^2_t$) of the returns, and finally, the square root is taken to obtain a forecast of the conditional volatility ($\sigma_t$), as undertaken by V-Lab. However, the square root of the variance is an inefficient estimate of the volatility (see [17]). Therefore, in the empirical work in this paper, we forecast volatility directly and not the variance.

Alternative approaches in papers such as [8, 16, 17] focus on the estimation of volatility (i.e., the standard deviation) of the investment’s returns and compute VaR forecasts as well as regularized risk forecasts based on generalized volatility models and neuro-volatility models. In [17], the authors proposed a data-driven generalized EWMA model based on sign correlation to estimate volatility directly and obtain optimal VaR forecasts. Neural networks (NNs) are one of the most common methods to approximate a multivariate nonlinear function. [16] applied a neuro-volatility model to forecast VaR with actual financial data. In [16], a data-driven neuro-volatility model is used to study the rolling neuro fuzzy forecasts of the Sharpe ratio (SR). In [8], regularized adaptive forecasts and computationally efficient forecasting algorithms for volatility, VaR, CVaR, and model risk are studied using various regularization methods such as ridge, lasso, and elastic net. In contrast to the related work cited above, this paper uses lagged values of the VIX as inputs to a neuro-volatility model (and other models as well).

We directly forecast the VIX itself (and its volatility). First, we use the historical time series of the VIX because it also contains the volatility risk premium, which is difficult to assess when using historical data on S&P returns, see [4] who find that GARCH models underforecast the VIX, and GARCH models display an inability to match option prices [3]. Second, [12] use options to forecast intraday values of the VIX. [13] argue that despite its theoretical foundation in option price theory, CBOE’s Volatility Index is prone to errors in deriving its value from options, may often be based on illiquid options, and has theoretical flaws. We therefore eschew options and use time-series data of the VIX itself.

The main findings of the paper are that nonlinear models (LSTMs) perform better with short-term forecasts of series like VIX, with non-stationary, long-memory data. Ensembles of linear and nonlinear models do well for longer horizon forecasts. For volatility of volatility, NNs (neuro-volatility models) do well, beating data-driven EWMA volatility models (though the latter have much better run times). The ensuing sections present VIX forecasting techniques and results, volatility of VIX forecasts, and conclusions.

2 MULTIPLE-STEP AHEAD VIX PRICE FORECASTS

The Cboe Volatility Index (VIX) represents the market’s option-implied near-term (30 day) forecast of S&P 500 index (SPX) volatility. It is based on the prices of SPX index options (calls and puts). The VIX began trading in March 2004 as

\[ \text{VIX} \]

\[ \text{https://vlab.stern.nyu.edu.} \]

Manuscript submitted to ACM
a futures contract, though it was first promulgated in 1993. Since then, the Cboe formula for the VIX has been changed to reflect a wider range of options. Because the VIX is derived from near-term call and put SPX option prices, it is a forward-looking measure of market volatility. Correspondingly, a forecast of the VIX is also a forecast of changes in option prices. As noted earlier, individual stocks returns and volatility are related to that of the market (through their stock beta), hence, the VIX is also an important ingredient in forecasting individual stock volatility.

We consider three different approaches as well as an ensemble approach as forecasts of VIX, described next.

### 2.1 Simple Moving Averages for VIX Forecasting

A simple moving average (SMA) calculates the average of a selected time period of prices, usually closing prices, by the number of periods in that range. A SMA is a technical indicator that can aid in determining if an asset price will continue or if it will reverse a bull or bear trend. We use SMA to forecast future VIX values. Assume we have \( n \) historical VIX closing prices \( P_1, \ldots, P_n \). In order to forecast the future \( D \)-day VIX, \( D = 1 \) number of the last price \( P_n \) will be imputed. Then we calculate the \( D \)-period SMA forecast \( \hat{P}_{t+1} \) as a function of \( P_t, \ldots, P_{t-(D-1)} \), which is a baseline forecast:

\[
\hat{P}_{t+1} = \text{SMA}_{t,D} = \frac{\sum_{i=t-D+1}^t P_i}{D}, \quad t = n, \ldots, D.
\]

### 2.2 LSTM for VIX Forecasting

The Long Short-Term Memory Network (LSTM network) is a type of Recurrent Neural Network (RNN). In a Traditional Neural Network, inputs and outputs are assumed to be independent of each other. However, LSTMs have loops inside them to have a memory of the previous computations and hence can handle the time series data. Unlike traditional RNNs, LSTMs do not usually encounter the vanishing gradient problem and exploding gradient problem. LSTM models are used to forecast non-stationary stock prices (usually modelled by geometric Brownian motion). The sample autocorrelation of the VIX series has a hyperbolic decay and hence it is non-stationary.

LSTM takes \( n \) historical VIX closing prices \( P_1, \ldots, P_n \) as the input and imputes \( D \) data points using the last observation \( P_n \). The LSTM outputs \( n - D \) one-step head points forecasts \( \hat{P}_{D+1}, \ldots, \hat{P}_n \) for \( P_{D+1}, \ldots, P_n \) and \( D \) future points forecasts \( \hat{P}_{n+1}, \ldots, \hat{P}_{n+D} \). Each forecast \( \hat{P}_{t+1} \) is a function of the past \( D \) data points \( P_t, \ldots, P_{t-(D-1)} \) obtained by LSTM.

### 2.3 The Prophet model for VIX Forecasting

Prophet is a time-series forecasting package developed at Facebook [15]. Their web page states that - “Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well.” We employ this model so as to capture the long-run dependence in the VIX data, to use a non-linear model, and to account for seasonality effects in case they are salient in the data.

Prophet has three additive components in its forecast model. (1) Trend (growth \( g \) over time, linear or nonlinear), (2) seasonality (\( s \), within year), cyclicity (across years), and (3) holidays (\( h \), irregular breaks). It is written as, for time series \( y \) (More extensive details are in the paper [15]):

\[
y(t) = g(t) + s(t) + h(t) + \epsilon_t,
\]

where \( \epsilon_t \) is the error term accounts for any unusual changes not accommodated by the model. Prophet re-frames the forecasting problem as a curve-fitting exercise.
Growth is modeled as 

(1) Growth is modeled as 

\[ g(t) = \frac{C}{1 + \exp(-k(t-m))} \]

where \( C \) is the saturation level, \( k \) is the slope, base growth rate, \( \frac{dk}{dt} > 0 \), \( m \) is the time offset, and \( C(t) \) can be made a function of time with exogenous analyst forecasts.

(2) Seasonality is modeled as a Fourier series. Here, \( P \): period, equal to 365 for annual, 7 for weekly. \( N \): components

of the Fourier series. \( s(t) = \sum_{n=1}^{N} a_n \cos \left( \frac{2\pi nt}{P} \right) + b_n \sin \left( \frac{2\pi nt}{P} \right) \).

(3) Holidays are modeled as follows. \( \kappa_i \) is change in forecast at time \( i \), written into a vector \( \kappa \). \( I(t) \) is an indicator vector of holiday dummies. Then \( h(t) = I(t) \cdot \kappa \).

### 2.4 Experiments

Time series data is different than non-sequential data when it comes to cross validation. We can create a cross validation sampling plan by offsetting the window used to select sequential sub-samples. In finance, this type of analysis is often called “backtesting”, which takes a time series and splits it into multiple uninterrupted sequences offset at various windows that can be tested for strategies on both current and past observations. We evenly distribute the data from 2016-01-04 to 2021-04-23 into 13 sets. Each window size is 504 days, and the skip span is 63 days. For each set, the last 21 days or 5 days are for the test (validation) set, and the rest are for the training set.

For each set, 21-day forecasts or 5-day ahead VIX forecasts are obtained using SMA, LSTM, or Prophet separately. An ensemble forecast is calculated as the average of the three (SMA, LSTM and Prophet) forecasts at each forecast time point. The forecast RMSE and MAE are calculated using the observed VIX from the test set and the forecasts. The results for the 21-day forecasts are summarized in Table 1, and the results for the 5-day forecasts are summarized in Table 2. The last two rows of Table 1 and 2 list the mean and standard deviation of forecast RMSE and MAE for each approach across data slices. For the 21-day forecast, there is no dominant approach for all the sets; therefore, the ensemble forecasts have the smallest forecast error (mean and standard deviation) of RMSE and MAE over all the sets. For the 5-day forecast, LSTM performs better than the other two approaches in general, as it has the smallest forecast errors over all the sets.

### 3 DATA-DRIVEN VIX VOLATILITY FORECASTS

Market participants also trade the volatility of the VIX itself, known as the VVIX. It is also known as the VIX of VIX. This measures the speed of change in market volatility and therefore corresponds to the volatility sensitivity of options (vega). VVIX may therefore be used to hedge volatility changes. Therefore, this paper assesses not only the forecasting of the VIX itself but also the volatility of VIX as these are complementary series. Based on the log returns of the VIX, we can obtain the risk forecasts such as volatility, VaR, and CVaR of VIX by using the sign correlation and identifying an appropriate \( t \) distribution. Moreover we can obtain intelligent probabilistic forecasts of the VIX.

Empirical studies in [17, 18] have shown that high-tech stock log returns follow heavy-tailed Student-\( t \) distribution with estimated degrees of freedom (d.f.) less than 4. Using the definition of sign correlation of a random variable given in [17], the sign correlation of the VIX log return \( \{ r_t, t = 1, \ldots, n \} \) with sample mean \( \bar{r} \) is defined as \( \hat{\rho}_r = \text{Corr}(r_t - \bar{r}, \text{sign}(r_t - \bar{r})) \). If \( r_t \) follows a Student’s \( t \) distribution with sample sign correlation \( \hat{\rho}_r \), the corresponding degrees of freedom (d.f.) \( \nu \) can be computed by solving the following nonlinear equation

\[ 2\sqrt{\nu - 2} = (\nu - 1)\hat{\rho}_r, \text{Beta} \left[ \frac{1}{2}, \frac{1}{2} \right] \]

The sign correlation plays an important role in finding the probabilistic forecasts of the VIX returns by determining an appropriate data-driven \( t \) distribution of the VIX returns. The following proposed volatility forecast models are based on the observed volatility of VIX log returns, which are defined as

\[ Z_t = \frac{|r_t - \bar{r}|}{2\hat{\rho}_r F(\bar{r})(1 - F(\bar{r}))} \]
where \( F(\bar{r}) \) is the cumulative distribution function (cdf) evaluated at the sample mean of VIX log returns. The observed volatility based on data-driven sign correlation incorporates skewness and non-normality.

Manuscript submitted to ACM
3.1 Data-Driven Exponential Weighted Moving Average (DD-EWMA) VIX Volatility Forecasts

For VIX log returns, summary statistics including absolute log returns and squared log returns show that time varying volatility models are more appropriate for the volatility estimate/forecast instead of the sample standard deviation (which does not take account the autocorrelations of the squared or absolute values and gives equal weights to the past values). Following [17], the DD-EWMA VIX volatility forecasting model is proposed as

\[ \hat{\sigma}_{t+1} = (1 - \alpha) \hat{\sigma}_t + \alpha Z_t, \quad 0 < \alpha < 1, \]  

where \( Z_t \) is the observed VIX volatility at time \( t \) and the sample sign correlation \( \hat{\rho}_r \) is used to identify an appropriate conditional t distribution of \( r_t \). Based on the past observations of VIX log returns, \( \hat{\rho}_r \) and \( Z_t \) are computed. The average of the first \( l \) observed volatility \( Z_1, \ldots, Z_l \) is used as the initial smoothed value \( \sigma_0 \), and the one-step ahead forecast error sum of squares (FESS) is calculated as \( \sum_{t=1}^l (Z_t - \hat{\sigma}_{t-1})^2 \). This volatility model is data-driven in the sense that the optimal value \( \alpha^* \) of the tuning parameter \( \alpha \) is obtained by minimizing the FESS. Using \( \alpha^* \), the optimal \( \hat{\sigma}_t \) is calculated recursively using (2), and the last optimal value is used as the one-day-ahead DD-EWMA volatility forecast (DDVF).

3.2 Neuro and Prophet VIX Volatility Forecasts

A NN can approximate any real nonlinear function on a compact domain to any degree of accuracy. Most of NN models in finance involve stock prices as the inputs. In a neuro volatility model, the inputs are the observed volatility. [16] proposed and studied a data-driven neuro volatility model for stock returns. In this paper, the nnetar function from the R Package forecast is used to fit the neuro volatility model to calculate VIX neuro volatility forecasts (NVF). The one-step ahead VIX NVF is computed using inputs that are lagged values of the observed volatility \( Z_t \), based on the sample sign correlation of VIX log returns.

Recently there has been a growing interest in using Prophet (R/Python packages) to forecast non-stationary time series based on observed stock prices. In this paper, the driving idea, unlike the existing work, is that Prophet is used to obtain the one-step ahead VIX volatility forecast (PVF), using observed volatility \( Z_t \).

3.3 Experiments

VIX log returns from 2019-04-26 to 2021-04-23 are used to forecast the VIX volatility for 2021-04-26. Summary statistics show that the t distributions with d.f. equal to 3.9284 (sample sign correlation \( \hat{\rho}_r = 0.7041 \) less than 0.7979 (normal distribution)) are more appropriate to model VIX log returns, which has sample mean 0.0005, standard deviation 0.0871, skewness 1.3954 and excess kurtosis 4.7351. The absolute series \(|r_t|\) and the squared series \( r_t^2 \) are significantly autocorrelated (acf(|\( r_t \)| = 0.2310, acf(\( r_t^2 \)) = 0.1674), which indicates volatility clustering. Therefore, we model the conditional distribution of \( r_t \) as a t distribution with mean \( \mu = 0 \) and changing volatility \( \sigma_t \) for the selected period. The d.f. of the t distribution is determined by the sample sign correlation from data. t-GARCH only applies well only for d.f. greater than four, so is marginally rejected as a candidate.

Results of VIX DDVF, NVF, and PVF for 2021-04-26 are first summarized in Table 3. It follows from Table 3, the RMSE of the data-driven neuro volatility model is the smallest one, while the run time to compute DD-EWMA volatility forecasts is faster than that to compute neuro volatility forecasts and Prophet volatility forecasts.

Let \( f(x) \) be the density function of the conditional distribution of log returns \( r_t \), and \( F^{-1}(p) \) be the inverse of the cdf of \( r_t \) evaluated at the tail probability \( p \). If the VIX log return follows a t distribution with d.f. \( v \), VaR and CVaR forecasts
with tail probability $p$ can be further calculated by the following equations:

$$\text{VaR}_p^R = -1000\hat{\sigma}_t \sqrt{\frac{v - 2}{v} \frac{1}{\nu} \left( 1 - F^{-1}(p, \nu) \right)}$$

$$\text{CVaR}_p^R = 1000\hat{\sigma}_t \sqrt{\frac{v - 2}{v} \left( \frac{f(F^{-1}(p, \nu))}{p} \left( \frac{v + (F^{-1}(p, \nu))^2}{v - 1} \right) \right)}$$

Forecasting the risk measures VaR$^R_p$ and CVaR$^R_p$ is equivalent to forecasting related functions of volatility and identifying an appropriate distribution for portfolio returns. Table 3 also summarizes the probabilistic VIX VaR and CVaR forecasts evaluated with tail probability $p = 0.05$ for 2021-04-26.

### Table 3. One-day-ahead volatility forecasts

<table>
<thead>
<tr>
<th>Day</th>
<th>DDVF RMSE</th>
<th>Time</th>
<th>NVF RMSE</th>
<th>Time</th>
<th>PVF RMSE</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2021-04-26</td>
<td>0.0746</td>
<td>0.0866</td>
<td>0.0030</td>
<td>0.0880</td>
<td>0.0662</td>
<td>1.9660</td>
</tr>
<tr>
<td>VaR$^{0.05}$</td>
<td>112.0622</td>
<td>169.1258</td>
<td>132.1736</td>
<td>199.4783</td>
<td>82.7612</td>
<td>124.9043</td>
</tr>
<tr>
<td>CVaR$^{0.05}$</td>
<td>VaR$^{0.05}$</td>
<td>CVaR$^{0.05}$</td>
<td>VaR$^{0.05}$</td>
<td>CVaR$^{0.05}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A rolling window approach is also applied to compare the rolling one-step ahead VIX volatility, VaR and CVaR forecasts using DD-EWMA, neuro volatility and Prophet. The selected data from 2019-04-26 - 2021-05-21 covers 1335 days of VIX log returns, and we compose 21 overlapping rolling windows, each of window size 1315 to calculate a one-day ahead VIX volatility, VaR and CVaR forecasts from 2021-04-26 to 2021-05-24. There is strong evidence from Figure 1 that the VIX volatility is time varying. VaR and CVaR forecasts will have similar results. Comparing to the rolling sample standard deviation of VIX log return, rolling sample standard deviation underestimates or overestimates the future volatility, VaR and CVaR. Moreover, if we look at the VIX price fluctuation for the period from 2021-04-26 to 2021-05-24 (left plot of Fig. 1), PVF overestimates the risk before 2021-05-09 and underestimates it after 2021-05-09, while DDVF and NVF perform more accurately to reflect the changing volatility and risk.

![Fig. 1. Comparison of one-day-ahead VIX rolling volatility forecasts with historical volatility: 2021-04-26 - 2021-05-24](image.png)

### 4 CONCLUSIONS

This paper presents a comparison of index volatility (and volatility of index volatility) forecasting approaches using the time series of the VIX and (absolute) VIX log return. We examined VIX forecasts over different horizons and found
that LSTM models do best for short-horizon forecasts and an ensemble model of SMA, LSTM, and Prophet does better for longer horizons. NN and DD-EWMA models do better than Prophet for the volatility of VIX returns. For various reasons outlined above, (see also [13]), we eschewed the use of options data to forecast the VIX. We also did not use the squared index return time-series, because of the large evidence that models such as GARCH that are fit to these series do not accurately capture the risk premium component of volatility [4]. Our main goal in this work is to compare forecasting approaches for the VIX using its own time series. Of course, there is also the possibility of using individual stock data to predict the VIX series as the cross-correlations have been found to be informative in this regard [14]. We are not sure if these additional data features will change the qualitative nature of the comparison of forecasting approaches, but they might improve the accuracy of all methods, although there is the attendant risk of overfitting the models as well. Overall, there is evidence that nonlinear forecasting methods are predicated for non-stationary series like the VIX.

5 ACKNOWLEDGMENTS

REFERENCES


Manuscript submitted to ACM