Intelligent Probabilistic Forecasts of VIX and its Volatility using Machine Learning Methods

Aerambamoorthy Thavaneswaran
Department of Statistics
University of Manitoba
Winnipeg, Canada
Aerambamoorthy.Thavaneswaran@umanitoba.ca

You Liang
Department of Mathematics
Ryerson University
Toronto, Canada
you.liang@ryerson.ca

Sanjiv Das
Leavey School of Business
Santa Clara University
Santa Clara, USA
srdas@scu.edu

Ruppa K. Thulasiram
Department of Computer Science
University of Manitoba
Winnipeg, Canada
tulsi@cs.umanitoba.ca

Janakumar Bhanushali
Department of Mathematics
Ryerson University
Toronto, Canada
janakkumar.bhanushal@ryerson.ca

***Abstract*—The market focuses on the Cboe Volatility Index (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. 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. Market participants such as traders, asset managers, and risk managers, keenly watch the VIX index, and are interested in achieving accurate intelligent probabilistic forecasts of the VIX, and also of the realized volatility of individual stocks. These volatility forecasts are useful to options traders placing bets on the future volatility of individual stocks. This paper examines 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 long short-term memory (LSTM) models, simple moving average methods, data-driven neuro volatility techniques, and industry models like Prophet. Uniquely, we propose a novel VIX price interval forecasting model. The driving idea, unlike the existing VIX price forecasting models, is that the proposed novel LSTM interval forecasting method trains two LSTMs to obtain price forecasts and the forecast error volatility forecasts. All the proposed forecasting methods also 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.**

***Index Terms*—Long memory, non-stationarity, LSTM, Prophet, Recurrent Neural networks, VIX, forecasts, Cboe**

I. Introduction

In this paper, we examine various approaches to forecasting the Cboe Volatility Index (VIX), 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 [1] 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 forecast 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 mean absolute deviation (MAD).

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 seasonal autoregressive integrated moving average (SARIMA) impose stationarity and normality. Second, some, but not all models assume normality. The work in [3], [4] shows that LSTMs do better in these settings.
2) The sample autocorrelations of the VIX series have hyperbolic decay. For this reason, we include a long short-term memory (LSTM) model and also use Facebook's Prophet [5] model, both of which have design features that will capture non-linearity and non-stationarity by design.
3) The family of generalized autoregressive conditional heteroskedasticity (GARCH) models for conditional variance are a poor choice for volatility forecasting of VIX. Volatility forecasting using traditional time-series methods, such as GARCH, exponential GARCH (EGARCH), asymmetric power ARCH (APARCH), Glosten-Jagannathan-Runkle (GJR) and integrated GARCH (IGARCH), [6], 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 constrain

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. [7], [8] show that assuming non-normality in VIX forecasting models improves on the normal models used in [9].
5) Commonly used volatility forecast models are historical simulation (HS), moving average (MA), Normal GARCH, Student-t GARCH, and exponentially weighted moving average (EWMA) for squared continuously compounded returns [6], [8]–[11]. Forecasts of the conditional variance (σ_t^2) of the returns are obtained, and finally, the square root is taken to obtain a forecast of the conditional volatility (σ_t), as undertaken by V-Lab (https://vlab.stern.nyu.edu). However, the square root of the variance is an inefficient estimate of the volatility (see [12]). Therefore, in the empirical work of this paper, we forecast volatility directly and not the variance.

Alternative approaches in papers such as [12]–[14] 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 [12], 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. [13] applied a neuro-volatility model to forecast VaR with actual financial data. In [15], a data-driven neuro-volatility model is used to study the rolling neuro fuzzy forecasts of the Sharpe ratio (SR). In [14], 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 [9] who find that GARCH models underforecast the VIX and GARCH models display an inability to match option prices [10]. Second, [16] use options to forecast intraday values of the VIX. [17] argues that despite its theoretical foundation in option pricing 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 on 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.

II. Multiple-step ahead VIX Price Forecasts

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 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.

The price of the VIX at time t, $P(t)$, is determined by the stochastic differential equation (SDE)

$$dP(t) = \mu P(t)dt + \sigma P(t)dW(t). \quad (1)$$

We consider three different approaches as well as an ensemble approach as forecasts of non-stationary VIX price series, described next.

A. 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, we first calculate one-day ahead forecast as

$$\hat{P}_{n+1} = \frac{\sum_{i=n-D+1}^{n} P_i}{D}.$$

For forecasting two-step ahead, we use the one-step forecast $\hat{P}_{n+1}$ as an input, along with the historical data $P_{n-D+2}, \ldots, P_n$. This process proceeds until we have computed all the required forecasts $\hat{P}_{n+1}, \ldots, \hat{P}_{n+D}$.

B. LSTM for VIX Forecasting

LSTM 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.

In Fig. 1, each line carries an entire vector, from the output of one node to the inputs of others. The pink circles represent pointwise operations, like vector addition, while the yellow boxes are learned neural network layers. Lines merging denote concatenation, while a line forking denotes its content being copied and the copies going to different locations.

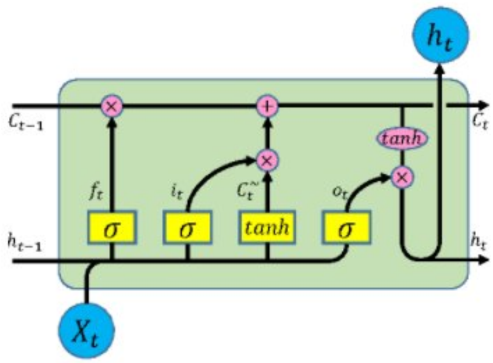

Fig. 1. LSTM cell

The LSTM cell contains the following components at each step t:

- Gating variables: forget gate f_t, input gate i_t, output gate o_t, which are NNs with sigmoid activation functions.
- Candidate cell state $\tilde{c}_t$ which is a NN with a tangent activation function.
- Cell and hidden states: memory state c_t and hidden state h_t.

Inputs to the LSTM cell at any step are X_t (current input) , h_{t-1} (previous hidden state) and c_{t-1} (previous memory state). Outputs from the LSTM cell are h_t (current hidden state) and c_t (current memory state).

LSTM models are used to forecast non-stationary stock prices (usually modelled by geometric Brownian motion (GBM)). The sample autocorrelation of the VIX series has a hyperbolic decay and hence it is non-stationary. We can divide the n historical VIX closing prices $P_1, \ldots, P_n$ into multiple inputs and output patterns. From the divided data in the input/output pattern, the LSTM model learn about how to use the inputs to forecast the output. Thus, P_t is one-step ahead output and its p lagged values $P_{t-1}, \ldots, P_{t-p}$ are the inputs at each step. The LSTM model produces one-step head point forecasts $\hat{P}_{p+1}, \ldots, \hat{P}_n$ for $P_{p+1}, \ldots, P_n$ and one-step ahead point forecast $\hat{P}_{n+1}$ using the inputs $P_{n-p+1}, \cdots P_n$. For forecasting two steps ahead, we use the one-step forecast $\hat{P}_{n+1}$ as an input, along with the historical data $P_{n-p+2}, \ldots, P_n$. This process proceeds until we have computed all the required forecasts $\hat{P}_{n+1}, \ldots, \hat{P}_{n+D}$.

C. The Prophet Model for VIX Forecasting

Prophet is a time-series forecasting package developed at Facebook [5]. 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), cyclicality (across years), and (3) holidays (h, irregular breaks). It is written as, for time series y (More extensive details are in the paper [5]):

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t, \tag{2}$$

where ϵ_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.

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, $\frac{\partial g}{\partial k} > 0$, 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. $s(t) = \sum_{n=1}^{N}\left[a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right)\right]$.
3) Holidays are modeled as follows. κ_i is change in forecast at time i, written into a vector κ. $I(t)$ is a indicator vector of holiday dummies. Then $h(t) = I(t) \cdot \kappa$.

D. Two LSTMs for VIX Interval Forecasting

In the following architecture, we will illustrate how to train two LSTMs to forecast D-day ahead VIX prices and construct VIX interval forecasts using the one-step ahead forecast errors of VIX prices.

- The first LSTM inputs n historical VIX closing prices $P_1, \ldots, P_n$ and outputs $n-p$ one-step head points forecasts $\hat{P}_{p+1}, \ldots, \hat{P}_n$ for $P_{p+1}, \ldots, P_n$ and D future points forecasts $\hat{P}_{n+1}, \ldots, \hat{P}_{n+D}$ as in section B.
- One-step ahead forecast errors are obtained from the first LSTM for P_t as $e_t = P_t - \hat{P}_t, t = p+1, \cdots, n$. We calculate the sample sign correlation $\hat{\rho}_e$ introduced in [12] of the forecast errors e_t and determine the corresponding normal or t distribution with appropriate degrees of freedom (d.f.) ν. The sign correlation of e_t with sample mean $\bar{e}$ is defined as

$$\hat{\rho}_e = \text{Corr}(e_t - \bar{e}, \text{sign}(e_t - \bar{e})). \tag{3}$$

If e_t follows a Student's t distribution with sample sign correlation $\hat{\rho}_e$ and finite variance, the d.f. ν can be computed by solving the following equation:

$$2\sqrt{\nu-2} = (\nu-1)\hat{\rho}_e \text{Beta}\left[\frac{\nu}{2}, \frac{1}{2}\right]. \tag{4}$$

- The sign correlation plays an important role in finding the probabilistic forecasts of the VIX prices by determining an appropriate data-driven t distribution of one-step ahead forecast errors. Moreover, one-step ahead forecast errors have time varying volatility. The volatility forecasts of e_t are then computed using inputs (the sample volatility of e_t) given by

$$Z_t = \frac{|e_t - \bar{e}|}{2\hat{\rho}_e F(\bar{e})(1 - F(\bar{e})},$$

where $F(\bar{e})$ is the cumulative distribution function (cdf) evaluated at the sample mean of e_t. The sample volatilities based on data-driven sign correlation incorporate skewness and non-normality as well.

- The second LSTM inputs the sample volatilities $Z_{p+1}, \ldots, Z_n$ and use q lagged values of the sample volatilities to forecast the one-step ahead volatility. The second LSTM model outputs one-step head volatilty forecasts of $e_{p+q+1}, \ldots, e_n$, as $\hat{\sigma}^e_{p+q+1}, \ldots, \hat{\sigma}^e_n$ and D-step ahead volatility forecasts $\hat{\sigma}^e_{n+1}, \ldots, \hat{\sigma}^e_{n+D}$.
- Probabilistic forecasts, for example, the D-step ahead interval forecasts, are obtained using point forecasts $\hat{P}_{n+1}, \ldots, \hat{P}_{n+D}$, volatility forecasts $\hat{\sigma}^e_{n+1}, \ldots, \hat{\sigma}^e_{D+n}$, and the t/normal distribution of the forecasting errors. The $100(1-\alpha)\%$ t forecast intervals for P_t's are given by

$$\hat{P}_t \pm t_{\alpha/2,\nu}\hat{\sigma}^e_t, t = n+1, \ldots, n+D. \quad (5)$$

For example, for $\nu = 4$ and $p = 0.05$, $t_{p/2,\nu} = 2.776$.

E. Experiment 1

1) VIX Price Forecasting: Time series data is different than non-sequential data when it comes to cross validation. We can create a cross validation sampling plan by splitting a time series into multiple uninterrupted re-samples that can be tested for strategies on both current and past observations. In finance, this type of analysis is often called "backtesting". We evenly distribute the data from 2016-01-04 to 2021-04-23 into 14 slices: 2016-01-04 - 2018-01-02, 2016-04-06 - 2018-04-05, 2016-07-07 - 2018-07-06, 2016-10-06 - 2018-10-05, 2017-01-09 - 2019-01-09, 2017-04-11 - 2019-04-11, 2017-07-13 - 2019-07-15, 2017-10-12 - 2019-10-14, 2018-01-16 - 2020-01-15, 2018-04-18 - 2020-04-17, 2018-07-19 - 2020-07-20, 2018-10-18 - 2020-10-19, 2019-01-23 - 2021-01-21 and 2019-04-25 - 2021-04-23. 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, either 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 errors (FE), RMSE and MAE, are calculated using the observed VIX from the test set and the forecasts. The LSTM model is implemented using Keras library and TensorFlow backend in R. The parameters of the model are presented in Table I.

TABLE I
INITIAL VALUES FOR THE LSTM FORECASTING MODEL

Parameters	Value
Numbers of Recurrent Units	128
Number of Dense Layer Units	64
Number of Lagged Observations	21
Optimizer	Adam
Batch Size	32
Epoch	50

TABLE II
CROSS VALIDATION FOR 21-DAY AHEAD VIX FORECASTS

Slice	SMA		LSTM		Prophet		Ensemble Model	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
1	**0.8170**	0.7373	0.9867	0.8607	1.5479	1.1209	0.8237	**0.7049**
2	3.5141	2.9891	5.6940	4.7097	**3.4152**	**2.5673**	3.7621	3.1676
3	**2.0109**	**1.6594**	2.2477	1.6641	4.8124	4.4595	2.0304	1.8327
4	1.0334	0.8991	1.0080	0.8915	1.8618	1.5617	**0.9619**	**0.6933**
5	6.6323	5.1417	**5.8308**	**4.2660**	7.1322	5.8096	6.4715	4.8574
6	**1.3014**	**1.1887**	2.7000	2.3167	6.1722	5.9494	3.2224	2.9911
7	2.5008	2.2211	2.7687	2.3502	**1.4190**	**1.1417**	2.1862	1.8133
8	3.1923	2.6825	4.7071	3.9768	**2.3748**	**1.8715**	2.8299	2.2500
9	1.1067	1.0016	2.0712	1.8071	1.8796	1.5726	**1.0205**	**0.8203**
10	14.0617	12.6757	20.2108	16.6686	23.3017	20.4362	**6.3865**	**4.8673**
11	3.2508	2.8514	**3.1906**	**2.8163**	13.0340	12.5472	5.6936	4.8711
12	**1.2418**	**1.0419**	2.0979	1.7513	5.3918	5.2627	2.6737	2.4126
13	1.9383	1.4778	2.6679	2.2600	1.7542	1.4073	**1.5063**	**1.2545**
14	4.1641	4.0772	4.3113	3.6526	2.0480	1.8407	**1.2150**	**1.0625**
14	1.4598	1.2988	1.1920	0.9597	1.1364	0.9600	**0.9030**	**0.6580**
FE mean	3.3404	2.9032	4.3209	3.5708	5.4389	4.8249	**2.9131**	**2.3999**
FE SD	3.4663	3.0974	4.8218	3.9559	6.0504	5.4689	**1.9858**	**1.5503**

TABLE III
CROSS VALIDATION FOR 5-DAY AHEAD VIX FORECASTS

Slice	SMA		LSTM		Prophet		Ensemble Model	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
1	**0.7031**	**0.5745**	0.7650	0.5896	1.5346	1.2372	0.6800	0.6012
2	2.5569	2.3066	1.8886	1.4857	2.5539	2.0768	**1.5260**	**1.1688**
3	1.8737	1.5117	1.6672	1.4670	1.6559	1.3085	**0.8890**	**0.6573**
4	1.4504	1.1928	**1.3769**	**1.0962**	2.1963	1.8598	1.5842	1.2004
5	4.9496	4.7456	**2.1768**	**2.1205**	2.6350	2.5093	3.0643	2.8558
6	**0.5609**	**0.5293**	1.0647	0.9365	5.8012	5.7085	2.3195	2.2945
7	**0.6840**	**0.6130**	1.6220	1.4365	1.6072	1.5017	1.2597	1.1164
8	2.1611	1.7143	**1.6793**	**1.5247**	2.7361	2.1171	1.8385	1.6421
9	**1.1870**	**1.1809**	3.0334	2.9191	1.9154	1.7078	1.3397	1.2080
10	4.7065	4.5199	**2.5105**	**2.0880**	5.4573	5.1612	4.1415	3.9230
11	3.1122	2.4832	**1.4856**	**1.2522**	13.1604	13.0672	5.6446	5.4183
12	**1.7100**	**1.3527**	2.0499	1.7649	4.8256	4.6212	2.7765	2.4831
13	**1.1147**	**0.9872**	1.1303	0.9222	2.0520	1.7030	1.3418	1.1472
14	1.4598	1.2988	1.1920	0.9597	1.1364	0.9600	**0.9030**	**0.6580**
FE mean	2.0164	1.7865	**1.6887**	**1.4688**	3.5191	3.2528	2.0935	1.8839
FE SD	1.3928	1.3381	**0.6100**	**0.6079**	3.1522	3.2104	1.4034	1.3978

The results for the 21-day point forecasts are summarized in Table II, and the results for the 5-day point forecasts are summarized in Table III. The last two rows of Table II and III list the mean and standard deviation (SD) of forecast RMSE and MAE for each approach across data slices. For the 21-day forecasts, there is no dominant approach for all the sets; therefore, the ensemble forecasts have the smallest forecast error (mean and SD) of RMSE and MAE over all the sets. For the 5-day forecasts, LSTM performs better than the other two approaches in general, as it has the smallest forecast errors over all the sets.

2) VIX Interval Forecasting: We investigate 5-day ($D = 5$) interval forecasts using slice 14. The first LSTM inputs the historical prices from 2019-04-25 - 2021-04-16, uses $p = 21$ lagged price values at each step, and outputs the one-step ahead forecasts from 2019-05-24 - 2021-04-16 and 5-day future price forecasts from 2021-04-19 - 2021-04-23. The forecast errors e_t's (Fig. 2) can be calculated from the differences between the predicted prices and the true prices from 2019-05-31 - 2021-04-16. Summary statistics show that that the t distributions with d.f. equal to 2.7757 (sample sign correlation $\hat{\rho}_e = 0.6036$ less than 0.7979 (normal distribution)) is more appropriate to model e_t, which has sample mean -0.7096, standard deviation 2.4141, skewness 1.6387 and excess kurtosis 11.5032. e_t is not significantly autocorrelated (acf($|e_t|$ = -0.0363). However, the absolute series $|e_t|$ and the squared

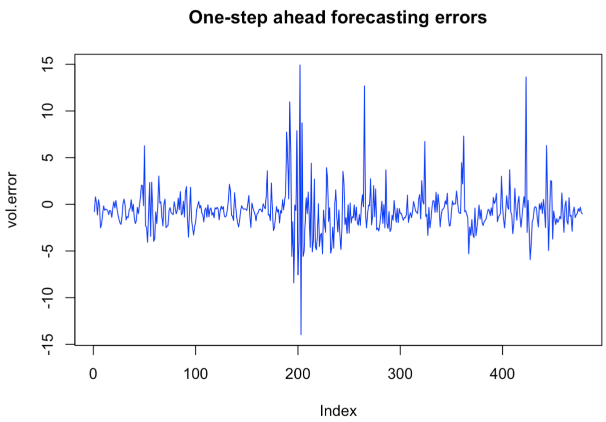

Fig. 2. One-step ahead forecast errors of prices

TABLE IV
5-DAY AHEAD VIX DATA-DRIVEN t INTERVAL FORECASTS: 2021-04-19 - 2021-04-23

Day	Price	$\hat{P}_t$	$\hat{\sigma}^e_t$	90% interval	95% interval	99% interval
1	17.290	17.246	1.421	(15.418, 19.074)	(14.742, 19.751)	(12.487, 22.006)
2	18.680	17.614	1.486	(15.702, 19.525)	(14.995, 20.232)	(12.637, 22.590)
3	17.500	17.997	1.571	(15.977, 20.017)	(15.230, 20.764)	(12.738, 23.256)
4	18.710	18.402	1.501	(16.471, 20.332)	(15.757, 21.046)	(13.376, 23.428)
5	17.330	18.849	1.494	(16.927, 20.771)	(16.216, 21.482)	(13.845, 23.853)

series e_t^2 are significantly autocorrelated (acf($|r_t|$ = 0.3543, acf(e_t^2) = 0.3523), which indicates volatility clustering.

Bollinger bands use the SMA as the middle line and use two standard deviations as the bandwidth assuming the normality of stochastic prices. However, LSTM forecasting errors follow a heavy-tailed t distribution with time varying volatility. The second LSTM uses the sample volatilities from 2019-05-24 - 2021-04-16 and uses $q = 5$ lagged values of the sample volatilities to forecast the one-step ahead volatility of forecast errors. Table IV summarizes the real VIX prices, point price forecasts, forecast error volatility forecasts and the interval price forecasts (calculated using (5)) from 2021-04-19 - 2021-04-23. The 90%, 95% and 99% forecast intervals contain the true prices for each day. It is also shown in Fig. 4 that the 95% forecast intervals contain the true prices over time. Moreover, the backtesting from 2019-06-03 - 2021-04-16 has sufficient evidence to demonstrate that the forecast intervals incorporate the time varying volatility of the forecast errors.

III. VIX Volatility Forecasts

The SDE (1) expresses the change in VIX price using a constant drift μ and volatility σ. Solving the above equation for $P(t)$ yields the solution:

$$P(t) = P(0)\exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W(t)\right\}.$$

Equivalently, we can express this equation as:

$$\log P(t) - \log P(0) = \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W(t).$$

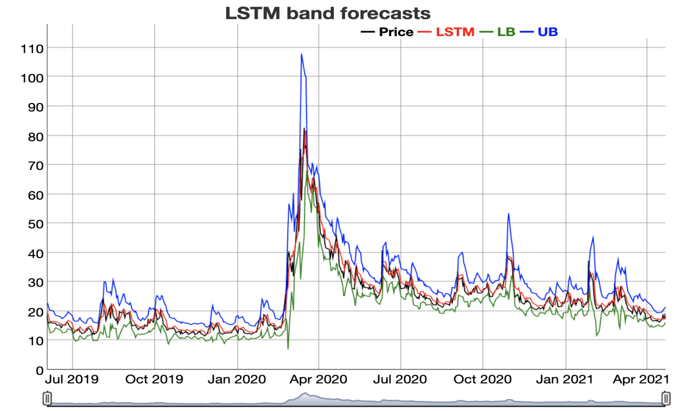

Fig. 3. 95% VIX forecast intervals using two LSTMs

GBM assumes the logarithmic change of the stock price to be a normally distributed random variable according to:

$$r_t = \log P_t - \log P_{t-1} = \left(\mu - \frac{1}{2}\sigma^2\right) + \epsilon_t, \epsilon \sim \mathrm{N}(0, \sigma^2)$$

The nonstationary VIX price is transformed to stationary (mean-reverting) log return series and the sample SD of the past data is traditionally used to estimate the volatility σ. 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, CVaR and MAD of VIX by using the sign correlation and identifying an appropriate t distribution. Moreover, we can obtain intelligent probabilistic forecasts of the VIX using the data-driven t distribution of VIX log returns.

In this section, three VIX volatility forecasting models with the implemented algorithms are described. Let the conditional mean and the conditional variance of the VIX log return r_t be

$$E(r_t|\mathcal{F}_{t-1}) = \mu_t, Var(r_t|\mathcal{F}_{t-1}) = \sigma_t^2, t = 1, \cdots, k,$$

respectively, where $\mathcal{F}_{t-1}$ is the past data up to time $t-1$. Empirical studies in [12], [18] have shown that high-tech stock log returns follow heavy-tailed Student-t distribution with estimated d.f. less than 4. The sign correlation of the VIX log return $\{r_t, t = 1, \ldots, n\}$ with sample mean $\bar{r}$ is defined as $\hat{\rho}_r = \mathrm{Corr}(r_t - \bar{r}, \mathrm{sign}(r_t - \bar{r}))$. If r_t follows a Student's t distribution with sample sign correlation $\hat{\rho}_P$ and finite variance, the corresponding d.f. ν can be computed by solving the following (4). The following proposed volatility forecast models are based on the sample 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 cdf evaluated at the sample mean of VIX log returns.

A. Data-Driven EWMA (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 [12], the DD-EWMA VIX volatility forecasting model is proposed as

$$\hat{\sigma}_{t+1} = (1-\alpha)\,\hat{\sigma}_t + \alpha Z_t, \quad 0 < \alpha < 1, \tag{6}$$

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 sample volatilities $Z_1, \ldots, Z_l$ is used as the initial smoothed value σ_0, and the one-step ahead forecast error sum of squares (FESS) is calculated as $\sum_{t=l+1}^{k}(Z_t - \hat{\sigma}_{t-1})^2$. This volatility model is data-driven in the sense that the optimal value α^* of the tuning parameter α is obtained by minimizing the FESS. Using α^*, the optimal $\hat{\sigma}_t$ is calculated recursively using (6), and the last optimal value is used as the one-day-ahead DD-EWMA volatility forecast (DDVF).

Algorithm 1 DD-EWMA VIX volatility forecasts

Require: Sample VIX volatilities $Z_t, t = 1, \ldots, k$
$S_0 \leftarrow \bar{Z}_l$ {Initial volatility forecast}
$\alpha \leftarrow (0.01, 0.3)\ by\ 0.01$ {Set a range for α}
$S_t \leftarrow \alpha * Z_t + (1-\alpha) * S_{t-1}, t = 1, \ldots, k$
$\alpha_{opt} \leftarrow \min_\alpha \sum_{t=l+1}^{k}(Z_t - S_{t-1})^2$
$S_t \leftarrow \alpha_{opt} * Z_t + (1-\alpha_{opt}) * S_{t-1}, t = 1, \ldots, k$
$\hat{\sigma}_{DD} \leftarrow S_k$ {Calculate one-step ahead DDVF}
return $\alpha_{opt}, \hat{\sigma}_{DD}$

B. Neuro 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 sample volatility. [13] 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 sample volatility Z_t, based on the sample sign correlation of VIX log returns.

C. Prophet VIX Volatility Forecasts

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 is that Prophet is used to obtain the one-step ahead VIX volatility forecast (PVF), using the sample volatility Z_t.

Algorithm 2 Neuro VIX volatility forecasts

Require: Sample VIX volatilities $Z_t, t = 1, \ldots, k$
$Vol.nnet \leftarrow nnetar(Z_t)$ {Fit a NN model}
$\hat{\sigma}_{NN} \leftarrow forecast(Vol.nnet, h = 1)$ {Compute one-step ahead NVF}
return $\hat{\sigma}_{NN}$

Algorithm 3 Neuro VIX volatility forecasts

Require: Sample VIX volatilities $Z_t, t = 1, \ldots, k$
$Vol.p \leftarrow prophet(Z_t)$ {Fit a Prophet model}
$\hat{\sigma}_{PP} \leftarrow predict(Vol.p, h = 1)$ {Compute one-step ahead PVF}
return $\hat{\sigma}_{PP}$

D. Experiment 2

We apply a rolling window approach to calculate the rolling one-step ahead VIX volatility forecasts using DD-EWMA, NN and Prophet. The selected data is the VIX log returns from 2019-04-25 -2021-05-21. We compose 21 overlapping rolling windows, each of window size 504 to calculate one-day-ahead VIX volatility forecasts from 2021-04-26 to 2021-05-24. For example, VIX log returns from 2019-05-23 to 2021-05-21 are used to forecast the VIX volatility for 2021-05-24. Summary statistics show that the t distribution with d.f. equal to 3.9284 (sample sign correlation $\hat{\rho}_r = 0.7041$ less than 0.7979 (normal distribution)) is more appropriate to model VIX log returns for this rolling window, which has sample mean 0.0006, standard deviation 0.0865, skewness 1.3590 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 σ_t for the selected period. The d.f. of the t distribution is determined by the sample sign correlation from the data. t-GARCH applies well only for d.f. greater than four so is marginally rejected as a candidate.

Results of rolling DDVF, NVF and PVF are summarized in Table V. It follows from Table V, accuracy determined by RMSE of all three methods is close, while the time using DD-EWMA volatility forecasts is faster than that using data-driven neuro volatility forecasts and Prophet volatility forecasts. An ensemble volatility forecast (ENVF) is calculated as the average of DDVF, NVF and PVF at each forecast time point. ENVF can obtain better predictive performance over time than that could be obtained from any of the constituent learning algorithms alone.

TABLE V

ONE-DAY-AHEAD ROLLING VOLATILITY FORECASTS FOR 21 TRADING DAYS: 2021-04-26 TO 2021-05-24

	DD-EWMA			NN			Prophet			Esemble
Day	DDVF	RMSE	Time	NVF	RMSE	Time	PVF	RMSE	Time	ENVF
1	0.075	0.087	0.003	0.074	0.070	2.029	0.060	0.082	1.394	0.069
2	0.067	0.087	0.002	0.070	0.070	1.770	0.068	0.082	1.486	0.068
3	0.058	0.087	0.002	0.044	0.068	1.987	0.073	0.082	1.248	0.058
4	0.053	0.087	0.003	0.065	0.069	1.719	0.080	0.082	1.270	0.066
5	0.048	0.087	0.002	0.062	0.070	1.696	0.117	0.082	1.338	0.076
6	0.053	0.087	0.002	0.077	0.069	1.715	0.070	0.082	1.587	0.067
7	0.048	0.087	0.003	0.071	0.068	1.762	0.077	0.082	1.251	0.066
8	0.055	0.087	0.002	0.075	0.068	1.732	0.083	0.082	1.282	0.071
9	0.050	0.087	0.003	0.073	0.069	1.733	0.083	0.082	1.350	0.069
10	0.051	0.087	0.002	0.070	0.069	1.734	0.018	0.082	1.510	0.047
11	0.065	0.087	0.002	0.099	0.071	1.710	0.077	0.082	1.286	0.080
12	0.092	0.087	0.002	0.059	0.069	1.720	0.089	0.081	1.308	0.080
13	0.101	0.087	0.002	0.121	0.075	1.650	0.101	0.082	1.308	0.108
14	0.143	0.088	0.003	0.114	0.069	1.723	0.104	0.082	1.525	0.120
15	0.166	0.088	0.002	0.235	0.069	1.760	0.131	0.082	1.253	0.177
16	0.195	0.088	0.002	0.160	0.070	1.747	0.085	0.082	1.297	0.147
17	0.168	0.088	0.003	0.142	0.070	1.727	0.096	0.082	1.289	0.135
18	0.156	0.088	0.002	0.096	0.072	1.732	0.098	0.082	1.550	0.116
19	0.138	0.088	0.003	0.090	0.072	1.748	0.099	0.082	1.241	0.109
20	0.132	0.088	0.002	0.047	0.073	1.718	0.041	0.082	1.291	0.073
21	0.116	0.088	0.002	0.086	0.072	1.738	0.061	0.082	1.317	0.088

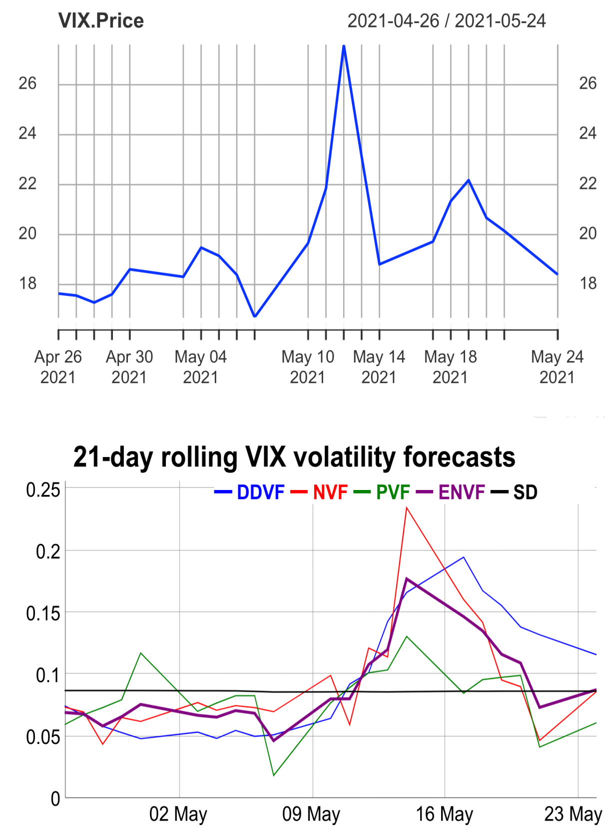

Fig. 4. Comparison of one-day-ahead VIX rolling volatility forecasts with historical volatility: 2021-04-26 - 2021-05-24

There is strong evidence in Fig. 4 that the VIX volatility is time varying. VaR, CVaR and MAD forecasts will have similar results. Compared to the rolling sample standard deviation of VIX log return, the 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. 4), PVF (green line) overestimates the risk before 2021-05-09 and underestimates it after 2021-05-09, while DDVF (blue line) and NVF (red line) perform more accurately to reflect the changing volatility and risk. ENVF (purple line) averages the three models to produce the final rolling forecasts.

Using volatility forecasts $\hat{\sigma}_t$ (DDVF, NVF and PVF), other intelligent probabilistic forecasts such as VaR, CVaR and MAD can be derived and calculated. 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. ν, VaR and CVaR forecasts with tail probability p can be further calculated by the following equations:

$$\mathrm{VaR}_t^p = -1000\hat{\sigma}_t\sqrt{\frac{\nu-2}{\nu}}F^{-1}(p,\nu),$$

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

The portfolio MAD forecast is computed as

$$\mathrm{MAD}_t = 2\hat{\rho}_r\hat{\sigma}_t\sqrt{F(\bar{r})(1-F(\bar{r}))}.$$

Forecasting the risk measures MAD_t, VaR_t^p and CVaR_t^p is equivalent to forecasting related functions of volatility and identifying an appropriate distribution for VIX returns.

The rolling window approach is also applied to compare the rolling one-step ahead VIX probabilistic forecasts using DD-EWMA, neuro volatility and Prophet. The selected data is VIX log returns from 2019-04-25 -2021-05-21. We construct 21 overlapping rolling windows, each of window size 504 to calculate a one-day ahead VIX VaR and CVaR with tail probability $p = 0.05$ in Table VI and MAD forecasts in Table VII from 2021-04-26 to 2021-05-24. Ensemble risk forecasts can obtain better predictive performance and will be selected in practice.

TABLE VI

ONE-DAY-AHEAD ROLLING RISK FORECASTS FOR 21 TRADING DAYS: 2021-04-26 TO 2021-05-24

	DD-EWMA		NN		Prophet		Ensemble	
Day	$\mathrm{VaR}^{0.05}$	$\mathrm{CVaR}^{0.05}$	$\mathrm{VaR}^{0.05}$	$\mathrm{CVaR}^{0.05}$	$\mathrm{VaR}^{0.05}$	$\mathrm{CVaR}^{0.05}$	$\mathrm{VaR}^{0.05}$	$\mathrm{CVaR}^{0.05}$
1	112.044	169.273	110.527	166.980	89.334	134.964	103.968	157.072
2	100.764	152.182	104.665	158.073	101.410	153.156	102.280	154.470
3	87.317	131.986	65.610	99.175	109.816	165.996	87.581	132.386
4	79.587	120.307	97.839	147.898	119.305	180.346	98.910	149.517
5	72.345	109.288	93.245	140.860	176.024	265.911	113.871	172.020
6	80.161	121.208	115.906	175.257	105.362	159.314	100.476	151.926
7	72.586	109.781	106.685	161.353	115.338	174.440	98.203	148.525
8	82.216	124.326	112.213	169.688	124.203	187.820	106.210	160.611
9	74.984	113.606	109.611	166.068	124.010	187.883	102.868	155.852
10	76.801	116.429	104.624	158.610	27.615	41.864	69.680	105.634
11	96.727	146.348	148.675	224.945	115.514	174.772	120.305	182.022
12	138.530	209.146	89.245	134.737	134.212	202.627	120.663	182.170
13	152.340	229.768	182.237	274.861	152.257	229.644	162.278	244.758
14	214.564	323.483	171.467	258.509	155.767	234.839	180.599	272.277
15	249.569	376.486	352.452	531.688	196.357	296.213	266.126	401.462
16	292.028	441.123	240.729	363.633	127.365	192.391	220.040	332.383
17	251.444	380.001	213.162	322.146	143.721	217.202	202.776	306.450
18	233.507	352.617	143.382	216.520	146.648	221.452	174.512	263.530
19	207.653	313.415	135.380	204.332	149.033	224.937	164.022	247.561
20	198.287	299.321	70.103	105.823	62.181	93.865	110.191	166.336
21	174.100	262.730	129.630	195.621	91.708	138.394	131.813	198.915

TABLE VII
ONE-DAY-AHEAD ROLLING MAD FORECASTS FOR 21 TRADING DAYS: 2021-04-26 TO 2021-05-24

Day	DD-EWMA	NN	Prophet	Ensemble
1	0.052	0.052	0.042	0.048
2	0.047	0.049	0.047	0.048
3	0.041	0.031	0.051	0.041
4	0.037	0.046	0.056	0.046
5	0.034	0.043	0.082	0.053
6	0.037	0.054	0.049	0.047
7	0.034	0.050	0.054	0.046
8	0.038	0.052	0.058	0.049
9	0.035	0.051	0.058	0.048
10	0.036	0.049	0.013	0.032
11	0.045	0.069	0.054	0.056
12	0.065	0.042	0.062	0.056
13	0.071	0.085	0.071	0.076
14	0.100	0.080	0.073	0.084
15	0.116	0.164	0.091	0.124
16	0.136	0.112	0.059	0.102
17	0.117	0.099	0.067	0.094
18	0.109	0.067	0.068	0.081
19	0.097	0.063	0.069	0.076
20	0.092	0.033	0.029	0.051
21	0.081	0.060	0.043	0.061

IV. CONCLUSIONS

We examined VIX price forecasts in section II over different horizons and found that LSTM models do best for short-horizon forecasts and a hybrid (ensemble) model of SMA, LSTM, and Prophet does better for longer horizons. Moreover, we proposed a novel method to train two LSTMs to obtain the interval forecasts of VIX prices. The driving idea, unlike the existing VIX price forecasting models using only one LSTM, is that the proposed approach trains one LSTM to forecast future prices and trains the other LSTM to forecast the volatility of the one-step ahead forecast errors. Overall, there is evidence that nonlinear forecasting methods are predicated for non-stationary series like the VIX as well as for cryptocurrencies, especially to construct future interval forecasts.

Moreover, Section III of this paper also presented a comparison of the VIX volatility (volatility of VIX returns) forecasting model using novel machine learning forecasting approaches based on the time series of the VIX and (absolute) VIX log return. NN and DD-EWMA models do better than the recently proposed industrial model Prophet for the volatility of VIX returns. There is evidence that models such as GARCH based on the squared index return time series do not accurately capture the risk premium component of volatility [9]. For various reasons outlined in [17]), we also eschewed the use of options data to forecast the VIX. Our main goal in this work is to compare novel forecasting approaches for the VIX using its own time series and the corresponding autocorrelations. 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 [19].

ACKNOWLEDGEMENT

The first and second authors acknowledge the Discovery grant from Natural Sciences and Engineering Research Council (NSERC).

REFERENCES

[1] A. Mashrur, W. Luo, N. A. Zaidi, and A. Robles-Kelly, "Machine Learning for Financial Risk Management: A Survey," *IEEE Access*, vol. 8, pp. 203 203–203 223, 2020, conference Name: IEEE Access.

[2] Cboe, "White Paper, Cboe Volatility Index," 2019. [Online]. Available: https://cdn.cboe.com/resources/vix/vixwhite.pdf

[3] J. Hosker, S. Djurdjevic, H. Nguyen, and R. Slater, "Improving VIX Futures Forecasts using Machine Learning Methods," *SMU Data Science Review*, vol. 1, no. 4, p. Article 6, Jan. 2019. [Online]. Available: https://scholar.smu.edu/datasciencereview/vol1/iss4/6

[4] M. Nunes, E. Gerding, F. McGroarty, and M. Niranjan, "Long short-term memory networks and laglasso for bond yield forecasting: Peeping inside the black box," arXiv.org, Tech. Rep. 2005.02217, May 2020, publication Title: Papers. [Online]. Available: https://ideas.repec.org/p/arx/papers/2005.02217.html

[5] S. J. Taylor and B. Letham, "Forecasting at scale," PeerJ Preprints, preprint, Sep. 2017. [Online]. Available: https://peerj.com/preprints/3190v2

[6] U. Majmudar and A. Banerjee, "VIX Forecasting," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 533583, Apr. 2004. [Online]. Available: https://papers.ssrn.com/abstract=533583

[7] Y. Jiang and E. Lazar, "Forecasting VIX Using Filtered Historical Simulation*," *Journal of Financial Econometrics*, no. nbaa041, Dec. 2020. [Online]. Available: https://doi.org/10.1093/jjfinec/nbaa041

[8] E. Berggren, "Are GARCH models necessary for Expected Shortfall?" 2017. [Online]. Available: http://lup.lub.lu.se/student-papers/record/8902182

[9] J. Hao and J. E. Zhang, "GARCH Option Pricing Models, the CBOE VIX, and Variance Risk Premium," *Journal of Financial Econometrics*, vol. 11, no. 3, pp. 556–580, Jun. 2013. [Online]. Available: https://doi.org/10.1093/jjfinec/nbs026

[10] C. Chorro and R. H. Fanirisoa Zazaravaka, "Discriminating Between GARCH Models for Option Pricing by Their Ability to Compute Accurate VIX Measures," *Journal of Financial Econometrics*, no. nbaa042, Jan. 2021. [Online]. Available: https://doi.org/10.1093/jjfinec/nbaa042

[11] M. V. Jon Daníelsson, Kevin R. James and I. Zer, "Model risk of risk models," *Journal of Financial Stability*, vol. 23, pp. 79–91, 2016.

[12] A. Thavaneswaran, A. Paseka, and J. Frank, "Generalized Value at Risk forecasting," *Communications in Statistics: Theory and Methods*, vol. 49, no. 20, pp. 4988–4995, 2020. [Online]. Available: https://doi.org/10.1080/03610926.2019.1610443

[13] A. Thavaneswaran, R. K. Thulasiram, Z. Zhu, M. E. Hoque, and N. Ravishanker, "Fuzzy value-at-risk forecasts using a novel data-driven neuro volatility predictive model," in *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, vol. 2, 2019, pp. 221–226.

[14] Y. Liang, A. Thavaneswaran, Z. Zhu, R. K. Thulasiram, and M. E. Hoque, "Data-Driven Adaptive Regularized Risk Forecasting," in *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, Jul. 2020, pp. 1296–1301, iSSN: 0730-3157.

[15] A. Thavaneswaran, Y. Liang, Z. Zhu, and R. K. Thulasiram, "Novel Data Driven Fuzzy Algorithmic Volatility Forecasting Models with Applications to Algorithmic Trading," in *Proceedings of 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2020, pp. 1–8.

[16] J. Osterrieder, D. Kucharczyk, S. Rudolf, and D. Wittwer, "Neural networks and arbitrage in the VIX," *Digital Finance*, vol. 2, no. 1, pp. 97–115, Sep. 2020. [Online]. Available: https://doi.org/10.1007/s42521-020-00026-y

[17] J. Osterrieder, L. Vetter, and K. Röschli, "The VIX Volatility Index - A Very Thorough Look at It," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3311727, Jan. 2019. [Online]. Available: https://papers.ssrn.com/abstract=3311727

[18] Z. Zhu, "Dynamic Data Science Applications In Finance," M.S. in Statistics, University of Manitoba, Manitoba, Canada, 2020.

[19] Y. Qiu, T. Xie, J. Yu, and Q. Zhou, "Forecasting Equity Index Volatility by Measuring the Linkage among Component Stocks*," *Journal of Financial Econometrics*, no. nbaa005, Jan. 2021. [Online]. Available: https://doi.org/10.1093/jjfinec/nbaa005