# Neural networks and arbitrage in the VIX

## A deep learning approach for the VIX

**Daniel Kucharczyk · Joerg Osterrieder · Silas Rudolf · Daniel Wittwer**

Received: date / Accepted: date

**Abstract** The Chicago Board Options Exchange (CBOE) Volatility Index, often referred to as VIX Volatility Index (VIX), is considered by many market participants as a common measure of market risk and investors' sentiment. It is also sometimes called the fear index. In general, the VIX represents the market's expectation of the 30-day-ahead looking implied volatility obtained from real-time prices of options on the Standard & Poor's 500 Stock Market Index (S&P500).
Over the last few years, many claims about possible VIX manipulations have been brought up by market participants. The increased attention on the VIX has been revived again by unusual trading patterns, which were observed on the market, on the 5$^{\text{th}}$ of February and 18$^{\text{th}}$ of April, 2018.
While smaller deviations between implied and realized volatility are a well-known stylized fact of financial markets, large, time-varying differences are also frequently observed throughout the day. In theory, such large deviations might lead to arbitrage opportunities on the VIX market. However, it is hard to exploit as the potential replication strategy requires buying several hundred out-of-the-money (put and call) options on the S&P500. In addition, the

Financial support by the Swiss National Science Foundation within the project "Mathematics and Fintech - the next revolution in the digital transformation of the Finance industry" is gratefully acknowledged by the corresponding author.

D. Kucharczyk
E-mail: danielkucharczyk@yahoo.com

J. Osterrieder
E-mail: joerg.osterrieder@zhaw.ch

S. Rudolf
E-mail: silas.rudolf@students.zhaw.ch

D. Wittwer
E-mail: daniel.wittwer@students.zhaw.ch

Institute
ZHAW, Technikumstrasse 9, 8401 Winterthur, Switzerland

potential list of options used for building the replication strategy constantly changes due to underlying price movements, making it difficult to implement it in real-time. Finally, in most cases, the theoretical replication strategy involves high transaction costs which are driven by illiquid options.
This paper discusses a novel approach to replicating and predicting the VIX by using just a subset of the most liquid options. The presented approach is based on a recurrent neural network, more precisely on a long short-term memory (LSTM) model and it uses intraday data of S&P500 options and the VIX. The results can be used to find a much more cost-efficient way of replicating the VIX and exploiting any arbitrage opportunities. To the best of the authors' knowledge, this the first paper, that describes a new methodology on how to replicate the VIX (to potentially exploit arbitrage opportunities using VIX futures) and applies most recently developed machine learning models to intraday data of S&P500 options and the VIX. The presented results are supposed to shed more light on the ongoing discussions about possible market manipulations, help other investors to better understand the market and support regulators to investigate market inefficiencies.

**Keywords** VIX · SPX · Neural Network · LSTM · deep learning

## 1 Introduction

The VIX index has been subject to claims of manipulation over the last few years, see e.g. Griffin and Shams (2017) [12]. We will analyze intraday data for S&P 500 options to predict the VIX, and, using neural networks, to show how one can exploit potential arbitrage opportunities without having to buy and sell several hundred out-of-the-money put and call options, as described by the VIX methodology [7] .
On February 5, 2018, the VIX moved the most in a single day in the index's 25-year history. The VIX and the VIX futures deviated substantially from each other on that day, which was one of the motivations behind our analysis. Another anecdotal evidence, showing the impact of SPX option trades on the VIX, is April 18, 2018. Shortly after the monthly settlement auction that determines the price for VIX options and futures, the VIX spiked as much as eleven percent within one hour. A trade of 13,923 May puts on the S&P 500 with a strike price of 1200, worth roughly $2.1 million, took place just as markets opened at 9:30 a.m. [18]
The Chicago Board Options Exchange (CBOE) Volatility Index (VIX) is a mathematical calculation which is considered the most important benchmark for implied volatility on the US stock market. Generally, the VIX sheds light on how investors "feel" about the market, hence its nickname, the "fear gauge." Its design is such that it tries to approximate the 30-day implied volatility of at-the-money options on the S&P 500. Anderson, Bondarenko, and Gonzalez-Perez (2015) [1] demonstrate that the VIX index can exhibit deviations from true volatility due to the inclusion of illiquid options. The methodology we

apply throughout the paper is to use a long short-term memory (LSTM) network architecture to analyse the time-series of S&P 500 option quotes and predict the VIX. Artificial neural networks have seen a revival in the last few years, due to better mathematical techniques for backpropagation but also due to the enormous computing power that is nowadays available. Recurrent neural networks which are composed of LSTM units are simply referred to as an LSTM network in the following. LSTM was proposed in 1997 by Sepp Hochreiter and Jürgen Schmidhuber [13] and improved in 2000 by Felix Gers' team [10].

Based on research by Kumar and Seppi in 1992 [16] and Spatt in 2014 [22], the S&P500 options and the VIX are markets with features that might leave it open to manipulation: the SPX options market with illiquid instruments and high transaction costs facing a large and liquid VIX derivatives market.

Any mispricings in the VIX should be arbitraged away by trading the VIX. Hoewer, this is not directly possible. One has to fall back to using VIX futures as a proxy for the VIX, to use the S&P 500 options to replicate the VIX or to find similarily suitable proxies, such as a limited set of put and call options to approximately replicate the VIX. Our goal is to use a neural network to show how to replicate and predict the VIX. The results are two-fold: First, we show how one can train the neural network to replicate the VIX, without knowing about the theoretical formula, by just using option prices. Second, we can also train the network to always use the out-of-the-money options that are closest to the current forward of the S&P index. There is a substantial benefit when applying this approach. By just using a small subset of all options that go into the VIX calculation, we can replicate the VIX with high accuracy and also predict its future value, beating the trivial approach of using the last observation as a prediction for the future value.

The remainder of the paper is organized as follows: Section 2 describes the historical relevance of the VIX for financial markets and introduces artificial neural networks as the method of choice for our analysis. Section 3 provides a literature review of relevant studies that use deep learning to analyse financial data and an overview of academic literature related to the VIX. In Section 4 we describe the data. Section 5 gives more background on the VIX and its relation to the option market by analysing the VIX formula, an option replication strategy and VIX futures. The design of the neural network, the implementation of the LSTM model as well as the results are described in Section 6. Finally, Section 7 discusses both the impact of this research and potential future applications.

## 2 The VIX and Deep Learning

Here we give some background information on the VIX and the deep learning technology that we apply. Section 2.1 discusses the relationship between the VIX and the S&P 500 options, while the required background on neural networks which is needed to understand the deep learning architecture is provided in Section 2.2.

## 2.1 The CBOE Volatility Index

We provide a short overview of the historical evolution of a volatility index on the U.S. equity market. Additionally, the current CBOE methodology for the computation of the VIX is explained. In the sequel, we will denote by relevant options those options that are used in the calculation of the VIX based on the CBOE VIX White paper [7]. Published by the CBOE, this volatility index is calculated using a weighted sum of mid-quotes, on out-of-the-money put and call options of the S&P 500 with a maturity between 23 and 37 days. [7]. Typically the VIX ranges between 10 and 30 points, major economic events being the exceptions. It cannot be traded directly, but there are many derivatives on the index, including options and futures. While entering the VIX as the square-root of weighted averages of prices, the SPX options contain much more information than the index itself, leading naturally to the possibility that there are different volatility surfaces implying the same VIX. Conversely, the same implied volatility can be achieved by different weighting and averaging schemes of the option prices, a feature which we will exploit later when applying our deep learning methodology.

### *2.1.1 Historical evolution of the VIX index*

In 1987, Brenner and Galai first introduced the Sigma Index in an academic paper [4]: "Our volatility index, to be named Sigma Index, would be updated frequently and used as the underlying asset for futures and options... a volatility index would play the same role as the market index play for options and futures on the index". In 1992, The American Stock Exchange announced a feasibility study for a volatility index, proposed as "Sigma Index". "SI would be an underlying asset for futures and options that investors would use to hedge against the risk of volatility changes in the stock market." On January 19, 1993, the Chicago Board Options Exchange introduced the VIX. Developed by Robert Whaley, it was designed to measure the 30 days implied volatility of at-the-money (ATM) S&P 100 (OEX) option prices [23]. 10 years later, the CBOE, together with Goldman Sachs, developed further computational methodologies which involved changing the underlying OEX to the S&P 500 (SPX). Generally, using SPX options with more than 23 days and less than 37 days to expiration ensures that the VIX will always reflect an interpolation of two points along the S&P 500 volatility term structure [7].
Up until now, this new VIX has been based on the S&P 500 registered Index (SPXSM), the core index for U.S. equities and estimates expected volatility by averaging the weighted quotes of SPX put and call options over a wide range of strike prices. In 2004, the CBOE began to introduce futures and two years later, in 2006, presented its new product, VIX options.

In 2014 another improvement was made by including SPX weekly options (SPXW), expiring on Fridays, in the calculation. This inclusion intends to more precisely reflect the 30 days expected volatility of the S&P 500.

#### *2.1.2 How the VIX market works*

The VIX, in its current form and methodology, has been in existence since 2014 and cannot be traded directly, since, after all, it is just a mathematical formula. However, derivatives including futures and options directly reference the VIX. Moreover, there are exchange-traded products (ETFs and ETNs) that offer investors exposure to the VIX.
The most important VIX-based derivative instruments that are in existence, as of 2018, include:

- 2004, VIX futures contracts
- 2006, Exchange-listed VIX options
- 2009, VIX futures based ETNs and ETFs, such as the S&P 500 VIX Short-Term Futures ETN (NYSE: VXX) and the S&P 500 VIX Mid-Term Futures ETN (NYSE: VXZ)
- 2010, S&P 500 VIX ETF (LSE: VIXS)
- 2011, VIX Short-Term Futures ETF (NYSE: VIXY) and VIX Mid-Term Futures ETF (NYSE: VIXM)

The VIX is being disseminated every 15 seconds from 2:15 a.m. to 8:15 a.m. and from 8:30 a.m. until 3:15 p.m. central standard time. The final settlement value for VIX futures and options is a Special Opening Quotation (SOQ) of the VIX Index calculated using opening prices of constituent SPX and SPX Weekly options that expire 30 days after the relevant VIX expiration date. For example, the final settlement value for VIX derivatives expiring on November 21, 2018, will be calculated using SPX options that expire 30 days later on December 21, 2018[1]. The opening prices for SPX options used in the SOQ are determined by an automatic auction mechanism on CBOE options, which matches locked or inverted buy and sell orders and quotes resting on the electronic order book at the opening of trading [8]. Even though the SPXW options expire at 3:00 p.m., the calculation for the settlement value takes place at the same time as the SPX options (8.30 a.m.).

#### *2.1.3 The CBOE VIX formula explained*

CBOE uses the following formula for the calculation of the VIX [7]:

$$\sigma^2 = \frac{2}{T} \sum_{i=1}^{n} \frac{\Delta K_i}{K_i^2} e^{rT} Q(K_i) - \frac{1}{T} \left( \frac{F}{K_0} - 1 \right)^2 \tag{1}$$

[1] If this Friday is a holiday, options will expire one day earlier and those are then used during the SOQ.

T is the time to expiration, r the risk-free interest rate (based on U.S. Treasury yield curves for the expiration dates of relevant SPX options), $F$ the forward price of the S&P 500 index, $K_0$ is the first strike below the forward index level $F$ and $K_i$ is the strike price of the $\mathrm{i}^{\mathrm{th}}$ OTM option. The quote $Q(K_i)$ is the mid-point of the bid-ask prices of the option with strike $K_i$. More precisely, $T$ is defined as follows:

$$T = \frac{(M_{\text{Current day}} + M_{\text{Settlement day}} + M_{\text{Other days}})}{\text{Minutes in a year}} \tag{2}$$

where $\mathrm{M}_{\text{Current day}}$ denotes the minutes remaining until midnight of the current day, $\mathrm{M}_{\text{Settlement day}}$ are the minutes from midnight until 8:30 a.m. for standard SPX options and minutes from midnight until 3:00 p.m. for SPXW expirations, $\mathrm{M}_{\text{Other days}}$ are the total minutes in the days between the current day and the expiration day of the options[2]. $F$ is defined as

$$F = \text{Strike Price} + e^{rT} \cdot (\text{Call Price} - \text{Put Price})$$

Here it should be pointed out that all calculations of the VIX are computed for the near- and next-term options. The CBOE distinguishes near-term options with a remaining time between 23 and 30 days and next-term options with a remaining term between 31 and 37 days.

When selecting the OTM puts you work successively from $K_0$ to the lower strikes and exclude all options with a zero-bid. If two consecutive zero bids occur, all options with lower strikes are no longer considered. Knowing all these rules and parameters one can easily calculate $\sigma_1^2$ and $\sigma_2^2$, which are the near- and next-term components of the VIX. To obtain the VIX value one takes a weighted 30-day average of $\sigma_1^2$ and $\sigma_2^2$

$$\text{VIX} = 100 \cdot \sqrt{\left[T_1 \cdot \sigma_1^2 \cdot \left(\frac{N_{T_2} - N_{30}}{N_{T_2} - N_{T_1}}\right) + T_2 \cdot \sigma_2^2 \cdot \left(\frac{N_{30} - N_{T_1}}{N_{T_2} - N_{T_1}}\right)\right] \cdot \frac{N_{365}}{N_{30}}} \tag{3}$$

where

1. $T_1$ = Time to expiry (as a fraction of the total number of minutes in a year) of the near-term options
2. $T_2$ = Time to expiry (as a fraction of the total number of minutes in a year) of the next-term options
3. $N_{T_1}$ = number of minutes to settlement of the near-term options
4. $N_{T_2}$ = number of minutes to settlement of the next-term options
5. $N_{30}$ = number of minutes in 30 days $(43,200)$
6. $N_{365}$ = number of minutes in a 365-day year $(525,600)$

Derman et al. in 1999 [5] show how the VIX formula can be derived based on a Brownian motion process for the underlying, using Black-Scholes assumptions, by using Itô's Lemma and approximating an infinite number of option strikes by a finite sum. Then, using various Taylor approximations as well as appropriate integral approximations, one arrives at the final formula for the VIX.

[2] A day contains 1440 minutes which is 24 hours.

## 2.2 Neural networks and deep learning

In the following, we will give a short overview of the techniques that we will later apply to our data. We will restrict ourselves to just those aspects that are needed in our analysis.
Deep learning (also known as deep structured learning or hierarchical learning) is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms. Learning can be supervised, semi-supervised or unsupervised.

A recurrent neural network (RNN) is a class of artificial neural networks where connections between nodes form a directed graph along a sequence. This allows it to exhibit temporal dynamic behavior for a time sequence.
We will use RNNs that have long short-term memory (LSTM) units. LSTM networks are well-suited to classifying, processing and making predictions based on time series data since there can be lags of unknown duration between important events in a time series. Relative insensitivity to gap length is an advantage of LSTM over RNNs, hidden Markov models and other sequence learning methods in numerous applications. LSTM was proposed in 1997 by Sepp Hochreiter and Jürgen Schmidhuber [13] and improved in 2000 by Gers et al. [10].
The major challenges of deep learning methods arise from the task of choosing the "best" model architecture. Facing the lack of computing power to test all possible model structures on any given data set, it is crucial to rely on previous research, data set characteristics, and intuition to design a deep learning model. David Wolpert, Mathematician and Santa Fe Institute Professor, describes the machine learning "no free lunch theorem" as follows: "for any two learning algorithms A and B... there are just as many situations (appropriately weighted) in which algorithm A is superior to algorithm B as vice versa." [24]. It follows that there is no universal model structure or learning algorithm, meaning different model structures give more accurate results on different data sets and for different purposes. There is also no universal guide on how to design a model, so intuition and experience are imperative in model design. Further complexities arise from underfitting and overfitting problems and from the task of how to efficiently train a neural network.
It is because of that no free lunch theorem, that we have to invest a substantial amount of time and effort into the precise network architecture. Following the proposed structure in [9], we describe the most important aspects of our network. Those are Initialization, Activation function, Normalization, Regularization, Optimizer and the learning rate schedule.

*Initialization* For our LSTM, we need to initialize the weights for the linear transformation of the input, the weights for the recurrent state and the bias vector. For those, we use the Glorot/ Xavier uniform initializer [11], the orthogonal initializer (the weight vectors associated with the neurons in each layer are supposed to be orthogonal to each other) and zeros, respectively.

The Glorot initializer achieves a good compromise for our desired requirement that the signal flows properly in both directions: in the forward direction when making predictions, and in the reverse direction when backpropagating gradients. For this to happen, the authors argue that we need the variance of the outputs of each layer to be equal to the variance of its inputs, and we also need the gradients to have equal variance before and after flowing through a layer in the reverse direction.

*Activation function* An activation function is used to introduce non-linearity to a network. This allows the model to learn complex mappings from the available data, and thus the network becomes a universal approximator, whereas, a model which uses a linear function (i.e. no activation function) is unable to understand complicated data. A good activation function is an important aspect when backpropagating through the network to compute gradients.
In our case, the *tanh* activation function is used: The Hyperbolic Tangent (tanh):

$$\tanh(x) = \frac{2}{1+e^{-2x}} - 1 \tag{4}$$

is a very popular and widely used activation function. It compresses the input in the range $(-1, 1)$ and provides an output which is zero-centered.

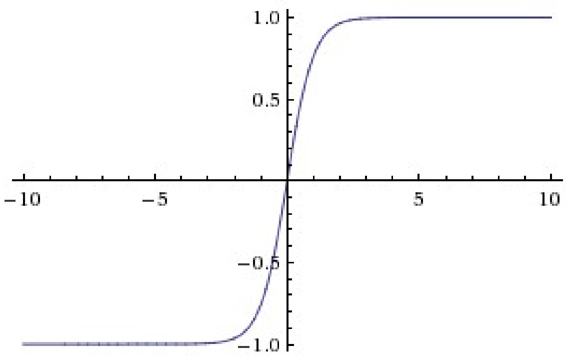


Fig. 1: The Hyperbolic Tangent (tanh) activation function

*Normalization* In a 2015 paper, Sergey Ioffe and Christian Szegedy [14] proposed a technique called Batch Normalization to address the vanishing/ exploding gradients problems, and more generally the problem that the distribution of each layers inputs changes during training, as the parameters of the previous layers change (which they call the Internal Covariate Shift problem). The technique consists in adding an operation in the model just before the activation function of each layer, simply zero-centering and normalizing the inputs, then scaling and shifting the result using two new parameters per layer (one for scaling, the other for shifting). In other words, this operation lets the model learn the optimal scale and mean of the inputs for each layer. We have decided to not use batch normalization since we are only using a batch size of ten in our application. Future research should focus on analysing this feature as well.

*Regularization* Deep neural networks typically have tens of thousands of parameters, sometimes even millions. With so many parameters, the network has an incredible amount of freedom and can fit a huge variety of complex datasets. But this great flexibility also means that it is prone to overfitting the training set.
As regularization technique, we have decided to use dropout at a rate of 0.1.

*Optimizer* Adaptive Moment Estimation (ADAM) is a method that computes adaptive learning rates for each parameter, in addition to storing an exponentially decaying average of past squared gradients, ADAM also keeps an exponentially decaying average of past gradients:

$$m_t = \beta_1 m_t - 1 + (1 - \beta_1) g_t \tag{5}$$

$$v_t = \beta_2 v_t - 1 + (1 - \beta_2) g_t^2 \tag{6}$$

$m_t$ and $v_t$ are estimates of the mean and the uncentered variance of the gradients. $g_t$ denotes the gradient, i.e. the vector of partial derivatives of $f_t$ evaluated at timestep t. $\beta_1$ and $\beta_2$ are hyper-parameters that control the exponential decay rates.
Since $m_t$ and $v_t$ are initialized as vectors of 0's they are biased towards zero. To counteract these biases, bias-corrected estimates are computed and used to update the parameters $\theta$ with the following ADAM update rule:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t \tag{7}$$

Summarizing, the benefits of ADAM consist of an adaptive learning rate and momentum for each parameter, as well as a non-diminishing learning rate. On the downside, it does not have the ability to "look ahead" before taking the next step like other optimizers, which include an approximation of $\theta_{t+1}$ in the calculation.

*Learning rate schedule* The Adam optimizer is an adaptive learning rate algorithm, therefore, we just need to decide on an initial learning rate. For the momentum decay hyperparameter, we use 0.9 and for the scaling decay hyperparameter, we use 0.999.

## 3 Literature Review

There have been many studies dedicated to investigating deep learning's applicability to financial problems involving classification and prediction. Most of those are forecasts of stock market returns. Olson and Mossman [20] in 2003 attempt to predict one-year-ahead stock returns for 2,352 Canadian companies using 61 accounting ratios as input values and reported that neural networks outperform traditional regression techniques. In 1993, Kryzanowski et al. [15] found that neural networks correctly classify 72% of the returns to predict one-year-ahead stock returns by using financial ratios and macroeconomic variables.

To predict one-day-ahead stock returns for the S&P500 constituents, Krauss et al. use deep neural networks, gradient-boosted trees and random forests. As a result, they show that combining the predictions of those three as an equal-weighted ensemble outperforms each individual model. Among each model, random forests outperform deep neural networks and gradient-boosted trees. Conversely, they stated that careful hyper-parameter optimization may still yield advantageous results for tuning intensive deep neural networks.

In 2016, Luca Di Persio and Oleksandr Honchar of the University of Verona completed a study that uses Artificial Neural Networks to predict stock market indices [6]. They experimented with many different architectures using Multi-layer Perceptron, Convolutional Neural Networks (CNN), and LSTM layers. Through a wavelet transformation (WT) technique, Periso and Oleksandr transformed their data before passing it through the CNN model, which produced the most accurate results out of all of the other models they used (including the CNN model without the transformed data). Another research team based in China similarly had success by combining WTs, stacked autoencoders (SAEs), and LSTM in a model for stock price forecasting [2]. Both of these studies highlight the importance of transforming the data in some way before passing it through a deep learning model in order to decrease noise.

The paper by Sepp Hochreiter and Jürgen Schmidhuber (1997) [13] is a comprehensive source on LSTM networks. In this study, the authors explain the mathematics behind why LSTM networks are able to solve complex problems that other networks are not. They also experiment with different types of datasets and compare LSTM's performance to other common networks. LSTMs and recurrent neural networks are still an area of intensive academic research and ongoing discussions. Recently, there has been a trend in handwritten text recognition with deep neural networks to replace 2D recurrent layers with 1D, and in some cases even completely remove the recurrent layers, relying on simple feed-forward convolutional only architectures. A more detailed discussion of that can be found in the 2018 paper of Moysset and Messina [19]. On the other hand those two authors show that 2D-LSTM networks still seem to provide the highest performances. The most important work on manipulation in the VIX was written by Griffin and Shams in 2017 [12]. They analyse market characteristics around the settlement of the VIX index in great details and show that volume spikes on S&P 500 index options at those times, but only for out-of-the-money options and more so for options with a higher and discontinuous influence on the VIX. Our goal is not to decide on which precise network architecture works best, we want to use an existing, widely used technique to apply it to a new datasets and to solve new questions arising from certain market characteristics of the option markets and the VIX. Derman et al. [5] have done the first comprehensive analysis and derivation of the price of volatility and variance swaps. They explain the properties and the theory of both variance and volatility swaps. They show how a variance swap can be theoretically replicated by a hedged portfolio of standard options with suitably chosen strikes, as long as stock prices evolve without jumps. For volatility swaps they show that those can be replicated by dynamically trading the more straightforward variance swap. Andersen et al. [1] demonstrate that the VIX Index has deviations from true volatility due to the inclusion of illiquid options. Futures and options on the VIX have a relatively large volume. The settlement value of those derivatives is calculated from a wide range of OTM put and call options with different exercise prices. A manipulator would have to influence exactly those prices of the lower-level OTM SPX options to influence the expiring upper-level VIX derivatives. The authors also show that fluctuations of illiquid OTM options lead to undesired variations of the VIX value. In 2017, Li [17] shows that the CBOE VIX methodology underestimates implied variance in general. The under-estimation increases as the forward index value moves higher and away from a strike price, peaks at the next strike, and resets to zero when passing the strike. He points out that a significant under-estimation can show up in related VIX indices such as the CBOE VVIX (the VIX of VIX) where fewer strikes are quoted. In 2018, Pimbley and Phillips [21] point out several aspects which show that the CBOE Volatility Index is prone to inadvertent and deliberate errors. They indicate several measures that can be taken to improve the index's accuracy and curtail its susceptibility to misuses and falsifications.

## 4 Intraday SPX Options and VIX Spot Data

Data is obtained directly from CBOE. The datasets contain SPX options as well as the VIX spot index, VIX futures and options on the VIX. We have intraday data for all data sets, for the S&P 500 options we have a one-minute granularity, the VIX itself is disseminated every 15 seconds and for VIX futures and options, we have a one-second granularity. The period of examination is the two-months period from January 2, 2018 until February 28, 2018, with the daily data available on trading days between 8:31 am CST until 3:15 pm CST. For the VIX index, only a particular subset of SPX options is used (see VIX white paper [7]. In our analysis, the filtering of the options is guided by the same methodology, but simplified in an appropriate way. We remove all options, at a given point in time, that do not satisfy all of the following criteria:

- Expiration date between 23 and 37 days in the future
- bid and ask greater than zero

- out-of-the money at the start of the given day

The features we use for our neural network are the log-returns of the mid-quotes of each option. This leads to an average of 500 available options per day between January 2018 and February 2018. The two-zero bid rule from the VIX methodology [7] is not considered in our analysis. Furthermore, we align the 15-seconds data for the VIX to the one-minute SPX option data.

## 5 VIX Highlights

Before we dive into the deep waters of neural networks to predict the VIX, we want to give a more detailed background on the VIX and the characteristics of the underlying options that are used to compute it. In Section 5.1, we will analyse the VIX formula and its two additive terms. Then, to motivate our approach of aiming to just use ten options, we show the number of options that are normally needed to fully replicate the VIX in Section 5.2. The events on February 5, 2018 are analyzed in Section 5.3 to show potential arbitrage opportunities in the VIX market.

### 5.1 Forward value in the VIX formula

In our analysis, we do not use the forward price, which enters the VIX via the second term of Equation (1). The justification can be seen in Figure 2. This term is very small compared to the actual VIX and can be neglected.

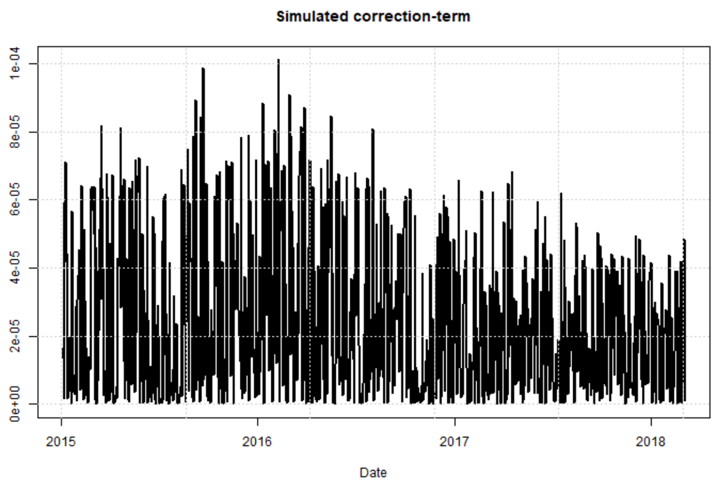


Fig. 2: Forward term of the VIX formula

## 5.2 Options in the VIX replication

We have also replicated the VIX based on the VIX white paper [7] to get a better understanding of what is needed to hedge it using options. In Figure 3 we see the evolution of the number of options for replicating the VIX between 2015 and 2018, which fluctuates between 200 and 450.

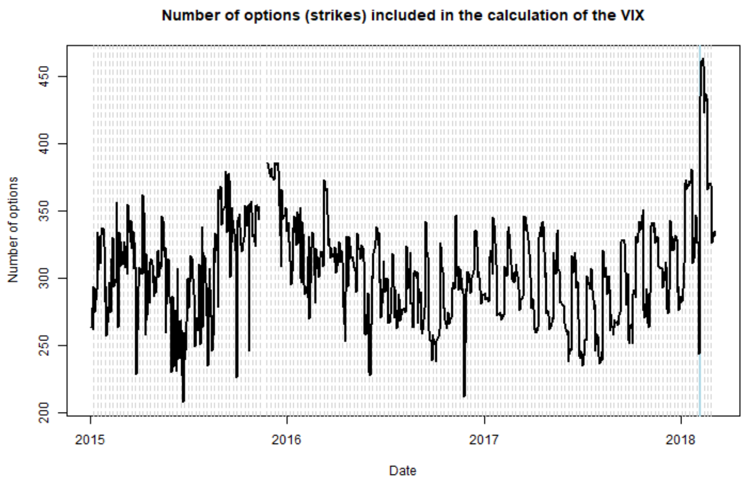


Fig. 3: Number of put and call options which are included in our VIX replication

Plotting the previous values separately by puts and calls, we see in Figure 4 that we need, on average, 100 put and 250 call options for the replication.

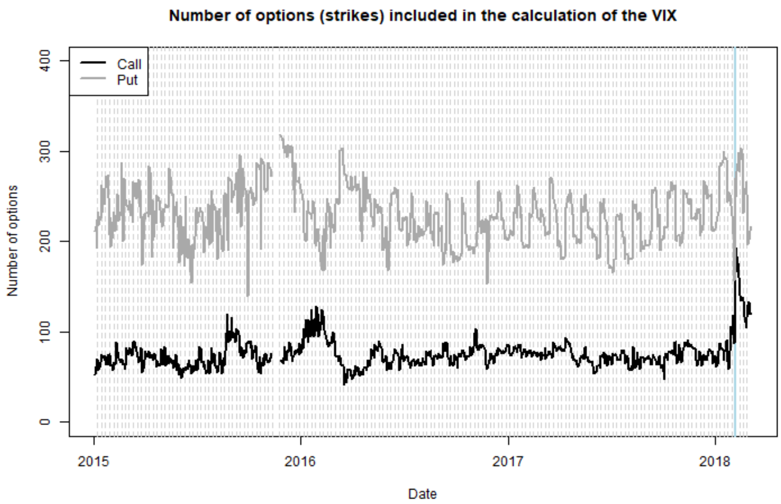


Fig. 4: Number of put and call options for the VIX replication, separated by puts and calls

### 5.3 The VIX and February 5, 2018

The events on February 5, 2018, when the VIX moved the most in a single day in the index's 25-year history, will be a strong motivation for our analysis, see Figure 5. On this day the VIX closed with 37.32 points, an increase of 20.01 points over the previous day, corresponding to an increase of 115% in one day. The extraordinary move coincided with a steep sell-off in the equity markets with the S&P 500 index falling by 4.1%. This event shocked the financial world and led to renewed accusations of market manipulation. On that day, we can observe a substantial deviation between the VIX and VIX futures. However, arbitraging away that difference is difficult, due to the sheer number of options that are theoretically needed, to fully replicate the VIX. Our approach later will simplify that task slightly, since we only need ten options to predict the VIX.


Fig. 5: VIX spot and mid quote on February 5, 2018

## 6 Using an LSTM network for predicting the VIX

We will use an LSTM network which is trained on SPX option quote data to predict the VIX value. For a given volatility surface, different ways of using the option quote data can be envisioned to replicate a given volatility surface. Therefore, on purpose, we do not use the VIX formula in our calculation, we simply use option quote data to train the network. The LSTM should be able to rediscover an appropriate way of combining this information, we do not want to impose any restrictions on it.

In Section 6.1 we describe the neural network architecture and in Section 6.2 we show the performance of the network for predicting the VIX.

### 6.1 Neural Network Architecture

The chosen architecture consists of one LSTM layer with 50 nodes, and one output layer with one node. The initialization is using the Glorot/ Xavier uniform initializer, the orthogonal initializer for the recurrent weights and zeros for the bias vector. For the activation function we use tanh. We do not use batch normalization since we only have a batch size of ten. Our data is normalized by computing log-returns of the prices. For regularization, we use a drop-out rate of 0.1. We have decided to use the ADAM optimizer with an initial learning rate of 0.9 for the momentum decay hyperparameter and 0.999 for the scaling decay hyperparameter .

The features we use are the log-returns of option prices of out-of-the-money put and call options. At the beginning of every day, we fix the set of options. The idea behind this is to simplify the process of actually trading those options. For the loss function, we are considering both the mean-squared error (MSE) for predicting the VIX returns and the categorical entropy for predicting up and down moves.

### 6.2 Predicting the VIX

On a normal day, about 350 options are needed for the replication of the VIX, see Figure 4, consisting of 250 put options and 100 call options. We will train our network on 10, 100, 200 options respectively, equally split between put and call options. Our training set is the intraday data in January 2018, the validation set is the data for February 2018, with a total of 1.68m and 1.52m observations, respectively (for 100 options).

In Figure 6 we show the MSE as a function of the number of epochs, for 10, 100 and 200 options.

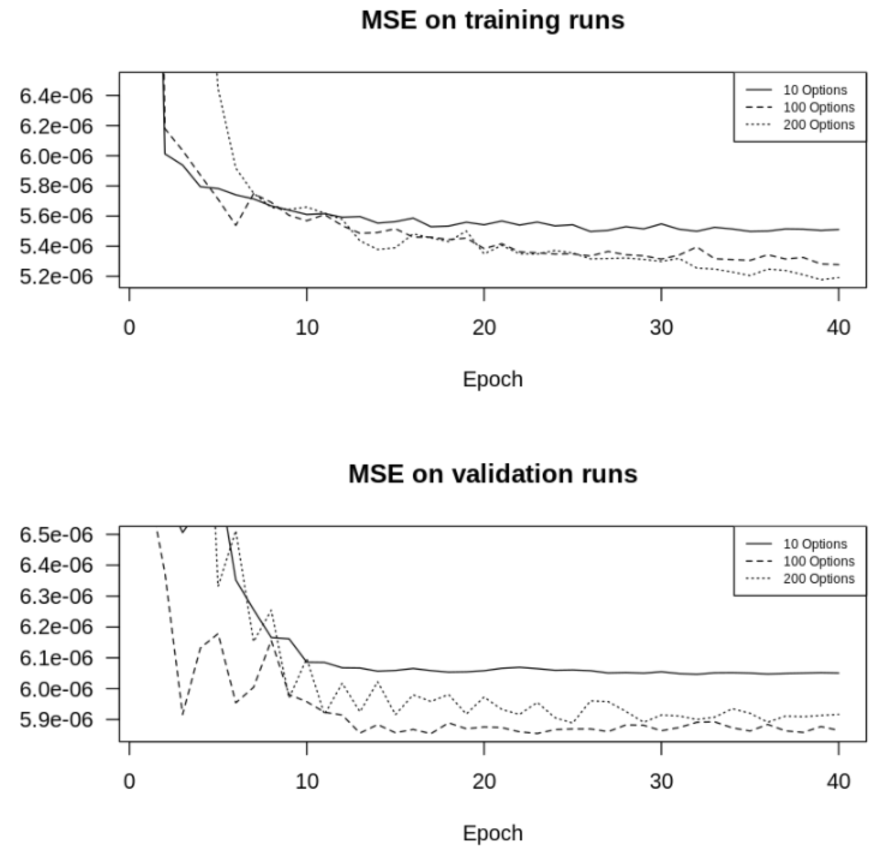


Fig. 6: MSE with one layer and 50 LSTM units, with 10, 100, 200 options

We summarize the MSE in Table 1.

Table 1: MSE with a varying number of options

| | 10 Options | 100 Options | 200 Options |
|---|---|---|---|
| Training data MSE | 5.51e-06 | 5.28e-06 | 5.19e-06 |
| Validation data MSE | 6.05e-06 | 5.86e-06 | 5.92e-06 |

For further visualization, the predicted VIX spot returns are compared with the actual values in the validation data set for ten options in Figure 7.

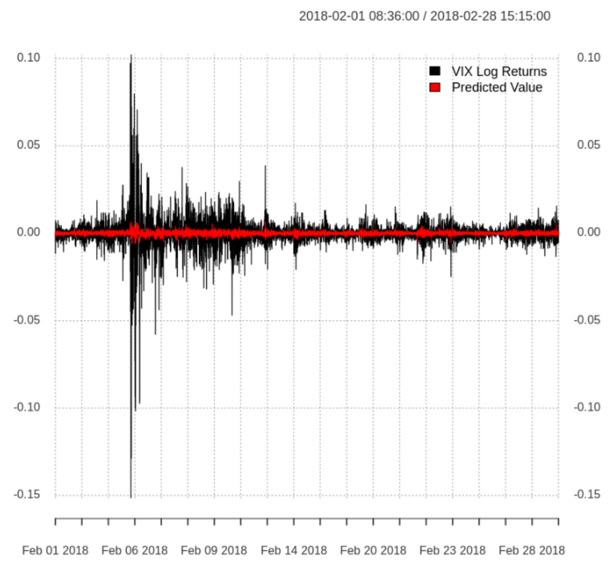


Fig. 7: Actual vs. predicted log returns using the validation data set

To judge how good the prediction is, we use the naive prediction which consists of just using the current VIX value as our forecast for the next time-step. The MSE for our benchmark is 5.43e-05 vs an MSE of 4.08e-05 for our prediction. Figure 8 shows the one minute ahead prediction of the VIX for one specific day, calculated as:

$$\hat{p}_i = p_{i-1} exp(\hat{r_i})$$

where $\hat{p}_i$ is the predicted price at time $i$, $p_{i-1}$ is the price at time $i-1$ and $\hat{r_i}$ is the predicted return at time $i$.


Fig. 8: VIX vs. one step ahead prediction

The MSE of our approach is better than the naive approach, but we need to shed more light on that result. We want to know how often our model predicts the correct direction of the price move. Using the categorical cross-entropy as loss function, with the sign of the option returns as input features and the sign of the VIX returns as output feature, we achieve an accuracy of 61.28% on the out-of-sample data, as can be seen in Table 2.

Table 2: Confusion matrix for our classification approach in February 2018

| | | Actual Signal | |
|---|---|---|---|
| | | -1 | 1 |
| Predicted Signal | -1 | 1252 | 721 |
| | 1 | 2222 | 3405 |

| PPV | NPV | Sensitivity | Specificity |
|---|---|---|---|
| 0.6051 | 0.6346 | 0.3604 | 0.8253 |

Figure 9 shows the improvement in the log loss as a function of the number of epochs. As expected, we get an improvement if we increase the number of options in our calculation. Remember that the VIX white paper [7] mandates that we use all out-of-the money options until we have two consecutive non-zero bids. Here, with only ten options we obtain good results, which makes it substantially easier to actually replicate the VIX.

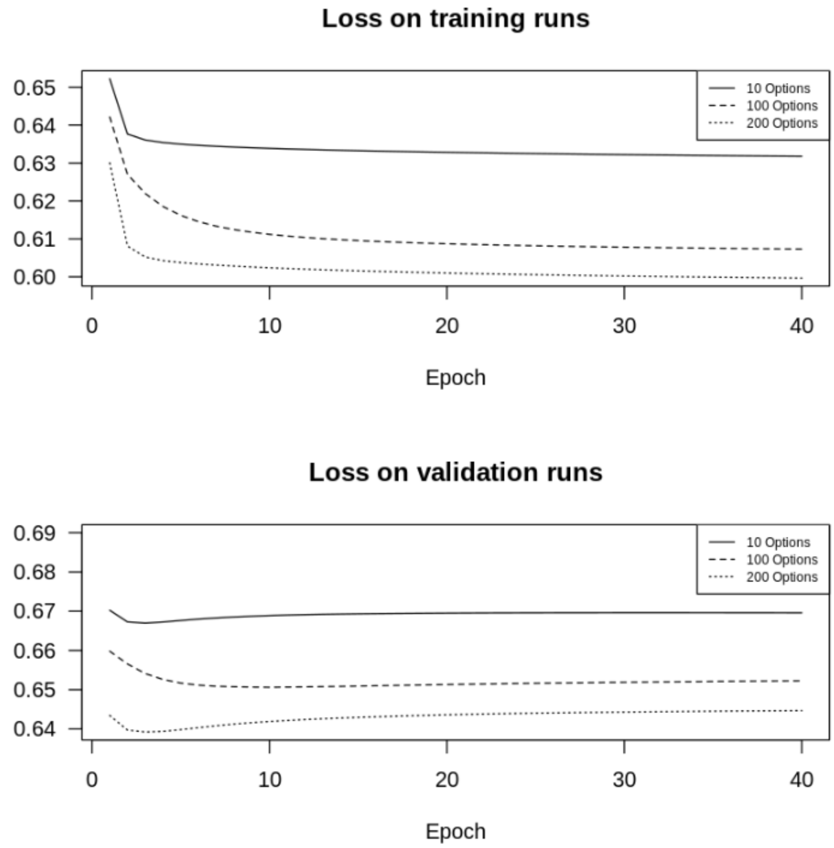


Fig. 9: Log Loss with 10, 100 and 200 options for our classification approach in February 2018

## 6.3 Random forests for the VIX

As a comparison to our deep learning approach, we have also used a more traditional machine learning approach, random forests. Our approach is based on Breiman's (2001) random forest implementation as described in [3]. Our random forest consists of 1000 trees, with three variables tried at each split, and input features consisting of the ten most important OTM options. Using the out-of-sample data for February 2018, we have recorded the results in Table 3, with an accuracy of 59.9%.

Table 3: Confusion matrix using random forests in February 2018

| | | Actual Signal | |
|---|---|---|---|
| | | -1 | 1 |
| Predicted | -1 | 1295 | 861 |
| Signal | 1 | 2179 | 3264 |

| PPV | NPV | Sensitivity | Specificity |
|---|---|---|---|
| 0.5997 | 0.6006 | 0.3728 | 0.7913 |

## 7 Conclusion and summary

To replicate the VIX using the official CBOE formula, one needs about 350 out-of-the money options at any point in time. We have shown that ten options (five call and five put options) are sufficient when used as input features for a neural network with one LSTM layer, to predict the VIX with an accuracy of 61.2%, which is slightly larger than using a random forest approach. Large deviations between VIX futures and the VIX arise on an intra-day scale. Using our methodology one might be in a better position to exploit any such arbitrage opportunities than is nowadays possible. Nevertheless, the option market is characterized by high transaction costs and low liquidity, which will still make it challenging to benefit from those differences between the VIX futures and its underlying. Further research in this area needs to focus on two aspects. Our approach, on purpose, was based on a simple LSTM to show the benefits of it, whereas future research can focus on refining the neural network architecture. The second aspect is to more precisely describe and analyse the arbitrage strategy that uses an appropriate subset of the S&P 500 options to replicate the VIX.

## References

1. Andersen, T.G., Bondarenko, O., Gonzalez-Perez, M.T.: Exploring Return Dynamics via Corridor Implied Volatility. The Review of Financial Studies **28**(10), 2902–2945 (2015)
2. Bao, W., Yue, J., Rao, Y.: A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PloS one **12**(7), e0180944 (2017)
3. Breiman, L.: Random Forests. Machine Learning **45(1)** (2001)
4. Brenner, M., Galai, D.: New Financial Instruments for Hedge Changes in Volatility. Financial Analysts Journal **45**(4), 61–65 (1989)
5. Demeterfi, K., Derman, E., Kamal, M., Zou, J.: More Than You Ever Wanted To Know About Volatility Swaps (1999)
6. Di Persio, L., Honchar, O.: Artificial neural networks approach to the forecast of stock market price movements. International Journal of Economics and Management Systems **1** (2016)
7. Exchange, C.B.O.: VIX White paper. https://www.cboe.com/micro/vix/vixwhite.pdf (2009). Accessed: 2018-05-26
8. Exchange, C.B.O.: Settlement Information for VIX Derivatives. http://cfe.cboe.com/cfe-products/vx-cboe-volatility-index-vix-futures/settlement-information-for-vix-derivatives (2018). Accessed: 2018-04-29
9. Géron, A.: Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. "O'Reilly Media, Inc." (2017)
10. Gers, F.A., Schmidhuber, J., Cummins, F.: Learning to Forget: Continual Prediction with LSTM. Neural Computation **12**(10), 2451–2471 (2000)
11. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. https://www.cboe.com/micro/vix/vixwhite.pdf (2010). Accessed: 2018-05-26
12. Griffin, J.M., Shams, A.: Manipulation in the VIX? The Review of Financial Studies **31**(4), 1377–1417 (2017)
13. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9**(8), 1735–1780 (1997)
14. Ioffe, S., Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. https://arxiv.org/abs/1502.03167 (2015)
15. Kryzanowski, L., Galler, M., Wright, D.: Using artificial neural networks to pick stocks. Financial Analysts Journal **49**(4), 21–27 (1993)
16. Kumar, P., Seppi, D.J.: Futures Manipulation with "Cash Settlement.". Journal of Finance **47**(4), 1485–502 (1992)
17. Li, A.: A Correction to the CBOE VIX Calculation Formula. https://papers.ssrn.com/sol3/papers.cfm?abstract$_id = 2991829(2017). Accessed : 2018-12-18$
18. Lu Wang, E.P., Kawa, L.: VIX Rigging Talk Erupts on Wall Street After Another Wild Swing. https://www.bloomberg.com/news/articles/2018-04-18/vix-rigging-talk-erupts-on-wall-street-after-another-wild-swing (2018). Accessed: 2018-11-27
19. Moysset, B., Messina, R.: Are 2D-LSTM really dead for offline text recognition? https://arxiv.org/abs/1811.10899 (2017). Accessed: 2018-12-18
20. Olson, D., Mossman, C.: Neural network forecasts of Canadian stock returns using accounting ratios. International Journal of Forecasting **19**(3), 453–465 (2003)
21. Pimbley, J., Phillips, G.: Fix the VIX: Reducing Manipulation in the Volatility Index. https://www.garp.org/!/risk-intelligence/all/all/a1Z1W000003fOTLUA2 (2018). Accessed: 2018-12-18
22. Spatt, C.: Security Market Manipulation. Annual Review of Financial Economics **6**(1), 405–418 (2014)
23. Whaley, R.E.: Derivatives on market volatility: Hedging tools long overdue. The Journal of Derivatives **1**(1), 71–84 (1993)
24. Wolpert, D.H., Macready, W.G.: No Free Lunch Theorems for Optimization. https://ti.arc.nasa.gov/m/profile/dhw/papers/78.pdf (1997). Accessed: 2018-12-18