Bitcoin Options Pricing Using LSTM-based Prediction Model and Blockchain Statistics

Lun Li
School of Computer and Information Technology
Beijing Jiaotong University
Beijing, China 100044
Email: lun.li@bjtu.edu.cn

Ali Arab
Department of Finance and Economics
Lubin School of Business
Pace University
New York, NY 10038
Email: aarab@pace.edu

Jiqiang Liu
(*Corresponding Author*)
School of Computer and Information Technology
Beijing Jiaotong University
Beijing, China 100044
Email: jqliu@bjtu.edu.cn

Jingxian Liu
School of Computer and Information Technology
Beijing Jiaotong University
Beijing, China 100044
Email: liujingxian@bjtu.edu.cn

Zhu Han
Department of Electrical and Computer Engineering
University of Houston
Houston, Texas 77004
Email: zhan2@uh.edu

***Abstract*-Although Bitcoin and other cryptocurrencies continue to attract attention, the inaccurate pricing of Bitcoin options has resulted in an inefficient crypto-derivative market. In this paper, we adopt a multiple input LSTM-based prediction model in conjunction with the Black-Scholes (BS) model to address this challenge in Bitcoin option pricing. We discuss the relationship between on/off-chain transactions and predict the implied price volatility of the next 30 days, which is the main factor in the BS model. First, we analyze the Blockchain statistics and social network trends of Bitcoin as inputs to our model, including the liveness of Blockchain wallet, the scale of active Blockchain nodes and the impact factor of Google, Reddit, Twitter to name a few. Next, we implement the LSTM-based prediction model and evaluate it in various historical window sizes and network parameters. Finally, we compare the performance with the baseline model without Blockchain statistics inputs. The experimental results show that the proposed multi-input LSTM-based prediction model provides the risk-informed pricing of the Bitcoin call options and our Blockchain statistics reduce the root-mean-square error (RMSE) by up to 46.2%.**

***Keywords*-Bitcoin, Blockchain, Prediction, Deep learning, LSTM, Option pricing.**

1. Introduction

Bitcoin is one of the earliest Blockchain-based peer-to-peer payment systems. Bitcoin has recently hit a very high market share. An increasing number of companies are supporting Bitcoin as a payment method to buy and sell products and services. Investors are becoming more enthusiastic about holding Bitcoin and its derivatives as an asset class. Bitcoin is a relatively complex asset for investors to acquire through on-chain peer-to-peer trading or mining. Therefore, they still need traditional exchanges for trading coins off the chain. A number of online marketplaces have been formed for trading Bitcoins and its financial derivatives including Bitcoin options. These markets are known as the off-chain Bitcoin markets. There are currently four markets that support the Bitcoin options [1], and the biggest Bitcoin option market, Deribit, [2] is trading 1,500 Bitcoin (i.e., BTC) per day. Moreover, Chicago Mercantile Exchange (CME) [3] announces to support Bitcoin futures to the market from 2018. Hence, acceptance is growing all over the world.

Bitcoin options traded in exchanges are European-style options, which means they cannot be exercised until expiry. The Black-Scholes (BS) [4] model is an effective way for pricing European options such as Bitcoin. However, one of the most important issues with the BS model is the inconsistency between the implied volatility in the BS model and the future observed market volatility in practice. Furthermore, unlike traditional stock markets, the Bitcoin market is still under an immature supervisory regime and affected significantly by the market sentiment online. Thus, historical volatility is usually extreme because of online marketing activities and investment sentiment. Many hedge funds and investors are unconfident in pricing Bitcoin options [5]. Although some existing work [6-9] considers using deep learning to predict the volatility of stocks, their models cannot be applied well to Bitcoin option pricing because Bitcoin on-chain transactions are publicly decentralized and the activities in the network influence the implied volatility of Bitcoin prices in the off-chain market. Thus, how to determine the price volatility for Bitcoin option pricing is still an open question.

In this paper, taking the unique characteristics of online Bitcoin trading and Blockchain statistics into consideration, we believe that it is more efficient to leverage a multiple inputs Long Short-Term Memory recurrent neural networks (i.e., LSTM network) in conjunction with the BS model for pricing Bitcoin options. Notably, the main contributions of this paper are as follows:

978-1-7281-4693-5/19/$31.00 ©2019 IEEE
DOI 10.1109/Blockchain.2019.00018

- First, we leverage a multiple-input LSTM neural network as the prediction model. The proposed model can predict the implied volatility of Bitcoin prices in the next 30 days and collaborate with the BS model for pricing the options.
- Next, we consider and collect the Bitcoin-related factors as the inputs to our predicted model. We collect the dataset including the traditional market statistics, Blockchain statistics, and social media trends. Traditional market statistics include the daily price, trading volume, and historical volatility; The Blockchain statistics analyzed in this paper includes wallet address liveness of Bitcoin exchanges in Blockchain and the active node scale in Blockchain. Social media trends include the impact factor of Google/Reddit/Twitter. In particular, analyzing Blockchain statistics is our main contribution in this paper.
- Finally, we conduct experiments and the evaluation of our models with the dataset from Aug. 8, 2015 to Nov. 17, 2018. The result shows that our model is adequately accurate for pricing 30-day Bitcoin call options and the analyzed Blockchain statistics have the obvious contributions to the prediction model, which reduce the root-mean-square error (RMSE) up to 46.2% compared to the baseline model without training with Blockchain statistics.

The remainder of this paper is organized as follows. Section 2 presents related work. Section 3 presents the fundamental preliminaries used in this paper. Section 4 describes data collecting and pre-processing. Section 5 presents the design of experiment and evaluations using the actual data in Bitcoin (BTC) exchanges. Finally, Section 6 concludes the paper.

2. Related Work

Ever since S. Nakamoto [10] proposed Bitcoin, Blockchain-based cryptocurrencies and Peer-to-Peer (i.e., P2P) payment systems are attracting various researchers. For the Blockchain research, researchers have explored various fields based on the tamper-proof and traceable features of Blockchain. N. Z. Aitzhan et al. [11] applied the Blockchain to the power trading system, F. Armknecht [12] proposed a real-time gross settlement system that uses the Blockchain-based network to handle bank transactions. L. Li [13] designed a Blockchain-based incentive mechanism in the vehicular network to encourage vehicles to send traffic information. Above Blockchain-based researches utilize the Blockchain's characteristics such as tamper-resistance and traceability, which shows a broad application prospect. In terms of the investment level, Bitcoin accounts for more than 50% of the market share from all crypto networks. With more and more support from various companies. It is expected that the advantages of Bitcoin will remain in the next few years. Therefore, many investors choose to invest in Bitcoin and its derivatives such as options.

For pricing the options, F. Black et al. [4] proposed the BS model, which is an effective way for pricing European-style options such as Bitcoin options. One of the most significant issues in their BS model is the implied volatility. Option pricing is biased if the model used volatility is inconsistent with future observed implied volatility. There exist researches that leverage machine learning models predict stock volatility. P. Tino et al. [14] proposed a Markov predictor model to trade on predictions of daily volatility. They combined sophisticated models for overcoming the nonstationary in the dataset, and they showed their model is as accurate as the Recurrent Neural Network (RNN) but with more simplicity. F. Wang [15] et al. researched on stock volatility using multi-kernel learning (MKL-ELM) based on extreme learning machine aiming to enhance the prediction performance. Their simulation result showed their model outperformed the Support Vector Machine (SVM). More recently, there has been additional research work based on improved LSTM and GARCH (i.e., Generalized Auto Regressive Conditional Heteroskedasticity) models for predicting the price movements of the stock market [6-9]. Particularly, K. Chen et al. [8] leveraged the LSTM-based model to predict the returns in the Chinese stock market, which was the very first work to demonstrate the volatility using LSTM. In recent work, H. Y. Kim et al. [7] designed a prediction model using GARCH-type models for the inputs of the LSTM-based model. They showed improved results for traditional LSTM in the prediction of stock volatility on the Korean stock index.

Aforementioned research work is constructed on the improvement of the algorithm for stock volatility. However, they cannot be performed directly on the volatility prediction of Bitcoin, due to the unique features for cryptocurrencies such as loose regulation, a public transaction ledger and daily fluctuations (more explanations are provided in Section 4). At present, only a few research work takes into account features of Bitcoin for the volatility or price prediction [16-18]. Colianni et al. [16] and E. Stenqvist et al. [18] used tweet sentiment analysis to make prediction specifically on Bitcoin's price volatility and investment decisions. A. H. Dyhrberg [17] analyzed the characteristic of Bitcoin volatility and leveraged the GARCH-based model to classify Bitcoin with gold and other asset classes. Our work focuses on the prediction of Bitcoin options using Blockchain statistics.

Different from the aforementioned research, we conduct an analysis of Bitcoin price volatility to make price volatility prediction and use the results for option pricing of Bitcoin in the BS model. We consider multiple inputs including the traditional marketing data such as daily trading volume and historical price and the unique factors specific to Bitcoin such as the Blockchain node scale and active frequency of wallet addresses.

3. Preliminaries

3.1. Black-Scholes Model

The Black-Scholes (a.k.a. Black-Scholes-Merton or BSM) model [4] is a pricing model to determine the fair price or theoretical value for options. In this model, Black et al. [4] assumed that the underlying price has constant implied volatility. Asset prices are continuous and subject to a lognormal distribution. The instantaneous log return of the asset price is a geometric Brownian motion. And the options cannot be exercised until expiry (i.e., European-style options), and non-divided value of the option price is, as follows:

$$C = SN(d_1) - N(d_2)Ke^{-rt},$$

$$d_1 = \frac{\ln\left(\frac{S}{K}\right) + (r + \frac{1}{2}\sigma^2)t}{\sigma\sqrt{t}},$$

$$d_2 = d_1 - \sigma\sqrt{t},$$

where,

- C is option price (i.e., call premium);
- S is current stock (i.e., Bitcoin) price;
- t is time until the options are exercised;
- $K$is option striking price;
- r is risk-free interest rate;
- N is cumulative standard normal distribution; and
- σ is the standard deviation (annual volatility).

$\mathrm{N}(d_2)$ is the probability of exercising the call. $\mathrm{N}(d_2)$ is the present value when the risk-free interest rate is r. As shown, the Black-Scholes formula is the difference between two simple binary options, which are asset-or-nothing call and cash-or-noting call.

3.2. Long Short-Term Memory

The LSTM network was proposed by Hochreiter et al. [19] to solve the Gradient explosion or vanishment issues in the traditional RNN model in deep learning. LSTM is leveraged with memory cells and forget/control gates to learn for either storing useful information or forgetting the unnecessary information.

Specifically, W_f, W_i, W_C and W_o are the weight matrices. b_f, b_C, b_i and b_o are the bias matrices. The forget gate layer in the unit is as follows:

$$f_t = \sigma\big(W_f \cdot [h_{t-1}, x_t] + b_f\big),$$

where σ is the activation function, h_{t-1} is the previously hidden state vector, and x_t is the current input.

The memory cell in the unit is as follows:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i),$$

$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C),$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t,$$

where the $tanh$ is the hyperbolic tangent function, C_t is the current memory cell vector, and C_{t-1} is the previous memory cell vector.

Finally, the output vector o_t and hidden state vector h_t is as follows:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o),$$

$$h_t = o_{t^*} \tan h(C_t).$$

TABLE I. DESCRIPTION OF COLLECTED DATA

Category	DETAIL	DESCRIPTION
Date	Date of Dataset	Daily data from 2015/08/08-2018/11/17
Market Statistic	BTC/USD price	Historical price form Yahoo Finance.
	Volume	Historical volume in major exchanges form Yahoo Finance.
Blockchain statistic	Address Liveness of Bitcoin exchanges	Daily withdraw and recharge transactions in major exchange's wallet addresses **(Algorithm 1)**
	Scale of Active Nodes	The amount of daily active nodes that be active in past 30 days. **(Algorithm 2)**
Social Media Trend	Reddit Post Trend	The daily trend of historical social media and Google search collected from API.
	Twitter Follower trend	
	Google Search trend	
Historical Volatility	Backward Volatility	The Backward volatility in past 30 days.
Realized Volatility	Realized Volatility	The realized volatility in next 30 days. **(for prediction)**

As the time series changes, the value of h changes frequently, and the value of C changes much slower than h to create the memory of *long short-term*. Each input vector x_t is determined by previous h_{t-1} and C_{t-1}. Therefore, LSTM performs better in longer sequences than normal RNN. Experimented with various datasets and flexible designs of network layers, LSTM has shown promising performance in robot control [20, 21] and sequence prediction [19].

4. Data Collection

In this section, we describe how we collect the data. The dataset for the experiment consists of five parts: general market statistics, Blockchain statistics, social network statistics, and volatility. Table I shows the description of the collected data. The details are presented in the following subsections.

4.1. General Market Statistics

The general market statistics consist of historical BTC/USD price and trading volume. First, we study the historical adjusted price of BTC/USD. Yahoo Finance [22] collects various financial indices and historical price for academic use. We collect the historical Bitcoin price in a data range from Aug. 8, 2015 to Nov. 17, 2018. We only keep the adjusted closing price since it represents the daily price when all transaction pairs are closed each day. Figure 1 illustrates the price movements during the

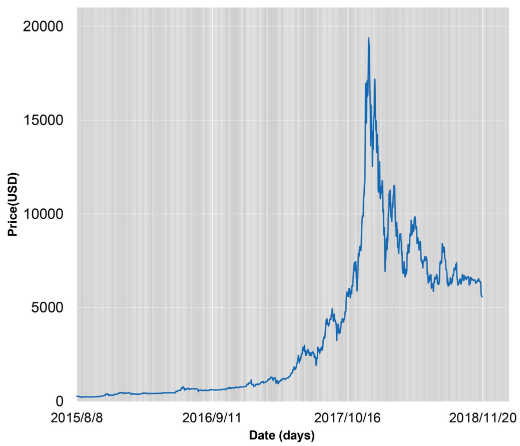

Figure 1. The Bitcoin Price (BTC-USD) Between Aug. 8, 2015 to Nov. 17, 2018.

aforementioned period. Next, we study the historical trading volume, which is a critical indicator of the market's movement and enthusiasm of the investors. The historical trading volume data from 79 cryptocurrency exchanges (including Binance, Coinbase, BitFinex, and Kraken, to name a few) are collected from CoinMarketCap [23].

4.2. Blockchain Statistics

In this paper, we study the impact of the Blockchain network on the off-chain market. Notably, our Blockchain statistics consist of the wallet address liveness and node scales. The volatility in the off-chain Bitcoin exchanges also reflects on-chain Blockchain records. We first explain the on-chain transaction and off-chain transaction and their relations in the following. Then, we propose our algorithm for processing the Blockchain statistics.

4.2.1. Explanation of On-chain and Off-chain Transaction

Since the Bitcoin system is based on Blockchain, a transaction in the Bitcoin system is called an on-chain transaction. Basically, there are three components in a Bitcoin transaction: the transaction's input ($txin$), the transaction's output ($txout$) and the transaction's ID ($txid$). In an on-chain transaction, there usually are multiple inputs and outputs so that all transactions on the Blockchain can form a large network. The coins and the balance in each account are calculated by codes instead of actual structures, which makes the Blockchain very reliable and tamper-proof. Specifically, each transaction's input is formed as follows:

$$txin^{(*)} \coloneqq \{\, txout\_prev, scriptsig\},$$

where $txout\_prev$ is the previous confirmed transaction's output, and $scriptsig$ is the unlock script to prove to the authority to spend the unspent transactions output.

Each transaction's output is formed as follows:

$$txout^{(*)}: = \{amount, scriptPubkey\},$$

TABLE II. TOP 6 ADDRESSES WITH HIGHEST BALANCE (MAR. 2019)

	OWNER	ADDRESS
1	Bittrex	385cR5DM96n1HvBDMzLHPYcw89fZAXULJP
2	Bitfinex	3D2oetdNuZUqQHPJmcMDDHYoqkyNVsFk9r
3	Bitstamp	3Nxwenay9Z8Lc9JBiywExpnEFiLp6Afp8v
4	Huobi	3Cbq7aT1tY8kMxWLbitaG7yT6bPbKChq64
5	Binance	3AweAnU1qYSUCJ5Hvy9DFEB7dVqUebZw5i
6		34xp4vRoCGJym3xR7yCVPFHoCNxv4Twseo

where $amount$ is the coin amount that needs to be spent on a certain address, and $scriptPubkey$ is the locking script that needs to be unlocked with required $scriptsig$. It is worth noting that $scriptPubkey$ usually contains the target wallet address of the payee, which we use to collect data for our LSTM-based model's input in this paper. The detailed information is shown in Section 5. The simplest definition of a functional on-chain transaction in the Bitcoin system is as follows:

$$tx \coloneqq (txid, txin, txout).$$

On the other hand, the off-chain transaction has few relations with Blockchain. Although Bitcoin is suitable as an anonymous dementalized payment system, due to the consensus algorithm *Poof-of-Work*, the confirmation of the on-chain transaction is not quick but relatively slow. According to [24], the entire Bitcoin system can only confirm less than 15,000 transactions per hour, compared to millions of transactions at NASDAQ. In addition, investors also need to join exchanges to contact the buyers and sellers similar to a traditional stock exchange. Therefore, currently the majority of Bitcoin online trading websites are off-chain trading, which means that the website requires investors to recharge their coins from their Bitcoin wallet addresses to the public wallet addresses. These addresses are public and owned by the exchanges. After confirming the transactions for recharging, their coins are displayed in the centralized account of this platform. The transactions on the market are not submitted to the Blockchain system. These transactions are called off-chain transactions. Only when investors choose to cash out their Bitcoin into their Bitcoin wallet, the coins cashed out will appear in the Bitcoin Blockchain.

4.2.2. Address Liveness of Bitcoin Exchanges

As described in the previous subsection, we analyze the address liveness in the Blockchain record. Specifically, we track the public wallet addresses of the top 10 Bitcoin exchange markets with the highest trading volume in the markets. We also analyze the address liveness by date. The address of each exchange can be found on their websites. For example, Table II lists the top six addresses with the highest balance, since the exchanges' wallets usually have a high

Algorithm 1: Statistic withdrawal and top-up transactions

Inputs: AllBlockchain record, ADDRLIST, date_range
Output: amount_wd, amount_topup
Start Procedure:
amount_wd=0, amount_topup=0
Block= getBlockchainRecord(date_range, AllBlockchain_record)
For i in len(Block):
  Tx_inblock[] = getTransaction(Block[i])
  For j in len(Tx_block):
    ***Check the item in ADDRLIST is in inputs' $scirptsig$:**
    If it is TRUE; amount_topup +=1;
    ***Check the item in ADDRLIST is in outputs' $scriptPubKey$:**
    If it is TRUE; amount_wd +=1;
Return amount_wd, amount_topup
End Procedure

Algorithm 2: Statistic of active Blockchain nodes

Inputs: AllBlockchain record, date_range
Output: DailyActiveNodesscale[]
Start Procedure:
  AddressListCount = 0
  For date in **range** date_range:
    DailyActiveBlock[] = getBlockchainRecord(range(date-30,date), AllBlockchain_record)
      For i in len(DailyActiveBlock):
        Tx_inblock[] = getTransaction(Block[i])
        For j in len(Tx_block):
        ***Check if the address in [$scirptsig$, $scriptPubKey$] is new:**
        If it is TRUE; AddressListCount+=1
  Return AddressListCount
End Procedure

balance of Bitcoin. The list of exchange public wallet addresses is collected.

Due to the transparency of Blockchain, we conduct analysis to track the exchanges' wallet addresses from public Blockchain historical records downloaded from the Bitcoin client. Algorithm 1 shows how to calculate the number of withdrawal and refill transactions related to the exchanges. As shown, the daily amount of withdrawal transactions and recharge transactions are being added and normalized.

4.2.3. Scale of Active Blockchain Nodes

In the Bitcoin network, the node with an address is considered a valid Bitcoin trading node. After transaction creation, transaction confirmation and block mining are depended on active Bitcoin nodes. Therefore, analyzing scale is an effective means of the volatility analysis for Bitcoin. Therefore, we analyze the public Blockchain records and record daily active nodes. Algorithm 2 shows how to measure the active nodes in each day. We only consider a node active when it appears in the past 30 days. If a new address appears in the record of the transaction, the corresponding node is considered to be active. Finally, the daily active node data is normalized and stored in our dataset.

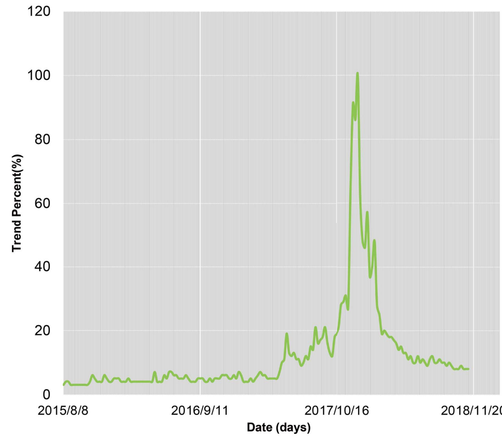

Figure 2. The Collected Social Media Trend from Aug. 8, 2015 to Nov. 17, 2018.

4.3. Social Media Trends.

As previously described, unlike the traditional stock markets, the Bitcoin market lacks a mature financial regulatory regime and supervisory framework, so it is easily affected by online market sentiment. A number of previous research work conducted Twitter sentiment analytics for price prediction [18]. The historical volatility is usually extreme because of online marketing activities and market sentiment.

In this study, we incorporate the market sentiments and social media trends in predicting implied volatility. In our dataset, we use the trends in social media as a simple measurement for marketing-heat. Daily Google trends of Bitcoin and Blockchain are collected from Google API [25]. Also, we collect the trends including the number of daily Reddit subscribers on Bitcoin and the number of the Twitter follower of 50 most popular accounts related to cryptocurrencies [26]. Specifically, the total trends are denoted as follows:

$$\frac{1}{3}\left(\frac{T_{Gi}-T_{Gmin}}{T_{Gmax}-T_{Gmin}}\right)+\frac{1}{3}\left(\frac{T_{Ti}-T_{Tmin}}{T_{Tmax}-T_{Tmin}}\right)+\frac{1}{3}\left(\frac{T_{Ri}-T_{Rmin}}{T_{Rmax}-T_{Rmin}}\right),$$

where T_{Gi}, T_{Ti} and T_{Ri} is the numbers of Google, Twitter and Reddit trends in target date. T_{Gmin}, T_{Tmin} and T_{Rmin} are the minimum values of each trend and T_{Gmax}, T_{Tmax} and T_{Rmax} are the maximum values of each trend. Figure 2 illustrates the social media trends we used from Aug. 8, 2015 to Nov. 17, 2018.

4.4. Volatility

For better performance, the backward volatility of past 30 days is included in the input. Moreover, the forward volatility of the next 30 days is calculated for training and testing the model and is also included in the dataset. Specifically, the 30-day backward annual volatility at day d (d is excluded) is calculated as follows:

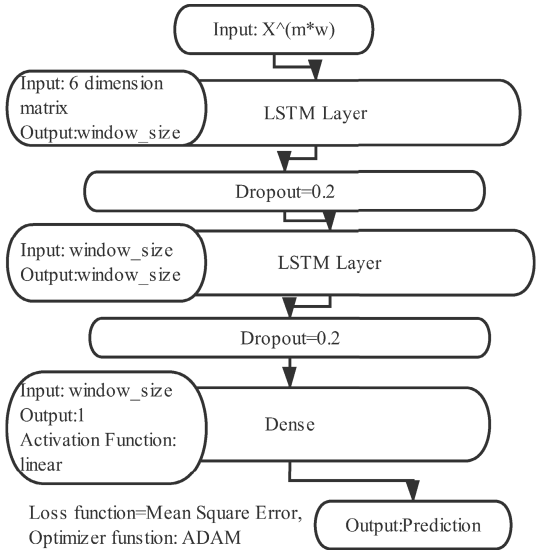

Figure 3. The Neural Network Structure of Proposed Prediction Model

$$BV_d = \sqrt{365}\sqrt{\frac{1}{n}\sum_{i1=d-n}^{d-1}(s_{i1} - \overline{s_d})^2},$$

where n is backward days (n=30), s_{i1} is the return rate from the $i1$ to $i1-1$, and $\overline{s_{i1}}$ is the standard deviation of $s_{i1},\ i1 \in \{1, \dots, n\}$. Note that, unlike the stock markets, we assume that there are usually 365 trading days for online Bitcoin exchanges.

Next, the 30-day forward annual implied volatility at day d (d is included) is calculated as follows:

$$FV_d = \sqrt{365}\sqrt{\frac{1}{n}\sum_{i2=d}^{d+n-1}(s_{i2} - \overline{s_{i2}})^2},$$

where n is forward days (n=30), s_{i2} is the return rate from the $i2$ to $i2-1$, $\overline{s_{i2}}$ is the standard deviation of $s_{i2},\ i2 \in \{1, \dots, n\}$

5. Experiment

In this section, we first denote our inputs and outputs and design our multiple input LSTM prediction model. Next, we use the collected dataset to train and test our model. Next, we compare the predicted FV_d with the actual FV_d for choosing the best window size in our experiment. Finally, we compare the predicted option price with the ideal option price using Root-Mean-Squared Error (RMSE) for evaluation.

5.1. Prediction and Evaluation Model

We use TensorFlow 1.9 [27] and Keras 2.2.4 [28] to build our prediction model and LSTM-based prediction model. The output of the network is simply predicted FV_d. For the inputs of our network, we use the input matrices with the variable window size of historical data and six dimensions of data described in Table I. Specifically, the input at day t denotes as follows:

$$Y_t = f(X^{m\times w}),$$

$$X^{m\times w} = [\boldsymbol{x_p}, \boldsymbol{x_{vol}}, \boldsymbol{x_{liv}}, \boldsymbol{x_{node}}, \boldsymbol{x_{soc}}, \boldsymbol{x_{BV}}],$$

where

a) $X^{m\times w}$ is the input matrices, m is the dimension and w is the window size;
b) $\boldsymbol{x_p} \in \mathbb{R}^{w\times 1}$ represents the historical price of Bitcoin. date range is $(t-w, t-1)$;
c) $\boldsymbol{x_{vol}} \in \mathbb{R}^{w\times 1}$ represents the historical volume in exchanges and date range is $(t-w, t-1)$;
d) $\boldsymbol{x_{liv}} \in \mathbb{R}^{w\times 1}$ is address liveness of Bitcoin exchanges and the date range is $(t-w, t-1)$;
e) $\boldsymbol{x_{node}} \in \mathbb{R}^{w\times 1}$ is the scale of active nodes and the date range is $(t-w, t-1)$;
f) $\boldsymbol{x_{soc}} \in \mathbb{R}^{w\times 1}$ is the data of social media trends and the date range is $(t-w, t-1)$;
g) $\boldsymbol{x_{BV}} \in \mathbb{R}^{w\times 1}$ is backward volatility and the date range is $(t-w, t-1)$;
h) Y_t is forward implied volatility at day t (for predictions); and
i) $f()$ is the LSTM prediction model;

The network structure is shown in Figure 3. We use LSTM layers and Dense layers for our prediction model. Also, we use full batch learning since it takes a shorter time when the training set is not very large. When choosing a small batch size, the results sometime can be more refined. However, it will be slow to converge and requires significant adjustments due to the small batch size. Therefore, we use the full batch leaning for the experiment.

The collected dataset is described in Section 4. We split the training set, validation set, and test set in 6:2:2. Next, we use the BS model with predicted FV_d to calculate the option price. Next, we calculate the option price using the BS model and actual FV_d. Finally, we evaluate our predicted option prices by RMSE. The option price is calculated as follows:

$$C = S * N(d_1) - N(d_2)Ke^{-rt},$$

$$d_1 = \frac{\ln\left(\frac{S}{K}\right) + (r + \frac{1}{2}\sigma^2)t}{\sigma\sqrt{t}},$$

$$d_2 = d_1 - \sigma\sqrt{t}.$$

The BS model is defined in Section 3. In our evaluation, first, the options are call options in the European style. Second, the risk-free rate is set to $r = 2.24\%$, which is Treasury rates in Mar. 2019 as described in [29]. Third, in current Bitcoin options exchanges, the exercise day usually is set to be between next half-month to a month. In our experiment, we predict the price in a month for the best usage. Therefore, time is $t = 1/12 = 0.833$. The RMSE is calculated as follows:

TABLE III. RMSE OF PREDICTED VOLATILITY/STRIKE PRICE IN TESTING DATASET

RMSE/Volatility			RMSE/10% 30-day Option Price			RMSE/50% 30-day Option Price		
Win. size	Epoch	RMSE[1]	Win. size	Epoch	RMSE[2]	Win. size	Epoch	RMSE[3]
80	500	0.188	50	700	82.012	50	700	38.794
50	700	0.190	40	100	85.327	80	1000	39.440
50	1000	0.197	50	1000	85.920	80	500	39.485
40	100	0.197	80	500	87.979	50	1000	39.999
70	300	0.204	10	700	89.747	70	300	41.072
60	100	0.207	70	100	90.158	50	400	42.077
50	900	0.208	70	300	90.383	10	700	42.198
10	700	0.209	50	900	90.543	60	1000	42.443
80	1000	0.210	70	200	91.141	30	700	42.491
30	200	0.211	10	900	93.069	70	100	42.569

*1 For the reference of RMSE, the average volatility in the test set is 0.29,
*2 For the reference of RMSE, the average price of 10% call options is 228.1,
*3 For the reference of RMSE, the average price of 50% call options is 41.81,
Experiment is conducted on TensorFlow 1.9 and keras 2.2.4.

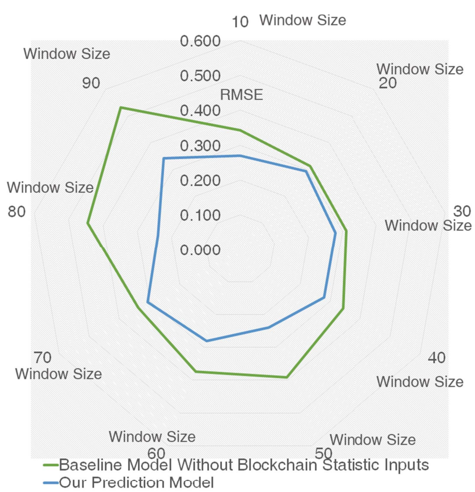

Figure 4. RMSE Comparison with Baseline Model which Excludes Blockchain Statistics Inputs

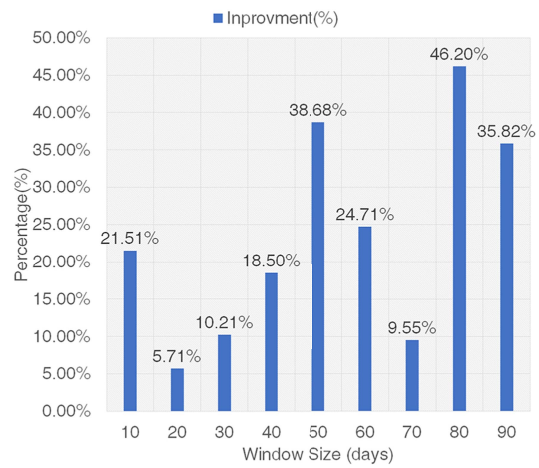

Figure 5. Percentage Reduction with Blockchain Statistics in Various of Window Size

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(C_{act}^{(i)} - C_{pre}^{(i)})^2}.$$

where i is the ith date of the training set. $C_{act}^{(i)}$ is the actual option price and $C_{pre}^{(i)}$ is the prediction of our model. We conduct the experiments 50 times in each setting and record the average RMSE.

5.2. Result Evaluation

We evaluate our model with various window size and training epoch in our predicted 30-day volatility, 30-day price with 10% call options and 30-day price with 50% call options. The partial results with the minimum RMSE are shown in Table III. As shown in the results, the best performance with minimum RMSE of implied volatility is 0.188 where the epoch in training is set to 500, and the window size is 80. The minimum RMSE of 10% call option price is 82.012; the minimum RMSE of 50% call option price is 38.794 where the window size is set to 50 and epoch when training is set to 700. In the experiment, the best window size is between 30 to 80. The RMSE will be more significant if we increase the window size over 80 since a large number of inputs lead to overfitting in LSTM layers.

Next, we compare the predicted volatility with the baseline input. The baseline model uses the same LSTM structure but trained without Blockchain statistics in inputs. The result is shown in Figure 4. With our analyzed Blockchain statistics, the RMSE of the predicted volatility is reduced effectively by up to 46.2% where the size of the window is 80, as shown in Figure 5. The minimum RMSE is 0.239 where the window size is 80. In general, the RMSE reduces more when it has a larger window size (up to 80), which means that Blockchain statistics in our prediction model have a better contribution to the model at the large window size.

6. Conclusion

We have proposed an LSTM-based prediction model for pricing Bitcoin call options. We collected our dataset including market statistics, Blockchain statistics, and social media trends. The market statistics include daily price, trading volume, and historical volatility. The Blockchain statistics include wallet address liveness of Bitcoin exchanges and the active node scale in Blockchain records. Social media trends include the impact factor of Google/Reddit/Twitter. The designed prediction model forecasts the implied volatility of the Bitcoin market in the next 30 days where it was used to predict the call option price using the Black-Scholes model. Furthermore, we conducted the experiment and the evaluated model performance. The result showed that the minimum RMSE of implied volatility was 0.188 with the window size of 80. Moreover, the RMSE for the price of 10% and 50% call options were minimum at window size 50. Finally, we compared the LSTM-based model without the inputs from Blockchain statistics. The results indicated that with Blockchain statistics, the RMSE was reduced by up to 46.2%. Therefore, Blockchain statistics had a significant contribution to the proposed prediction model.

Acknowledgments

The corresponding author is Jiqiang Liu. This work was supported in part by the Natural Science Foundation of China under Grants 61672092, in part by the Fundamental Research Funds for the Central Universities of China under Grants 2018JBZ103.

References

[1] BambouClub. "Trading Bitcoin Options at Deribit," Accessed at Mar. 20, 2019; https://hackernoon.com/trading-Bitcoin-options-at-deribit-part-1-155e9b7fc5dc.
[2] "Deribit.com," (Accessed at Mar. 20, 2019); https://www.deribit.com/.
[3] "CME Group to Launch Bitcoin Futures December 18, 2018" (Accessed at Mar. 20, 2019); https://news.bitcoin.com/cme-group-to-launch-bitcoin-futures-december-18/.
[4] F. Black, and M. Scholes, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy,* vol. 81, no. 3, pp. 637-654, 1973.
[5] N. Reiff, "Why Aren't Hedge Funds Interested in Bitcoin?," https://www.investopedia.com/news/why-arent-hedge-funds-interested-bitcoin/, [(Accessed at Mar. 20, 2019), 2017].
[6] C.-H. Tseng, S.-T. Cheng, Y.-H. Wang, and J.-T. Peng, "Artificial neural network model of the hybrid EGARCH volatility of the Taiwan stock index option prices," *Physica A: Statistical Mechanics and its Applications,* vol. 387, no. 13, pp. 3192-3200, May 2008.
[7] H. Y. Kim, and C. H. Won, "Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models," *Expert Systems with Applications,* vol. 103, pp. 25-37, Aug. 2018.
[8] K. Chen, Y. Zhou, and F. Dai, "A LSTM-based method for stock returns prediction: A case study of China stock market," in *Proc. IEEE International Conference on Big Data (Big Data)*, Santa Clara, CA., Nov. 2015, pp. 2823-2824.
[9] E. Haugom, S. Westgaard, P. B. Solibakke, and G. Lien, "Modelling day ahead Nord Pool forward price volatility: Realized volatility versus GARCH models," in *Proc. 7th International Conference on the European Energy Market*, Madrid, Spain, Jun. 2010.
[10] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," https://bitcoin.org/bitcoin.pdf, [Accessed: Oct. 2018, 2008].
[11] N. Z. Aitzhan, and D. Svetinovic, "Security and Privacy in Decentralized Energy Trading through Multi-signatures, Blockchain and Anonymous Messaging Streams," *IEEE Trans. Dependable Secure Comput.,* vol. 15, no. 5, pp. 840-852, Sep. 2018.
[12] F. Armknecht, G. O. Karame, A. Mandal, F. Youssef, and E. Zenner, "Ripple: Overview and Outlook," in *Proc. International Conference on Trust and Trustworthy Computing*, Heraklion, Crete, Greece, Aug. 2015.
[13] L. Li, J. Liu, L. Cheng, S. Qiu, W. Wang, X. Zhang, and Z. Zhang, "CreditCoin: A Privacy-Preserving Blockchain-Based Incentive Announcement Network for Communications of Smart Vehicles," *IEEE Trans. Intell. Transp. Syst,* vol. 19, no. 7, pp. 2204-2220, Jan. 2018.
[14] P. Tino, C. Schittenkopf, and G. Dorffner, "Financial Volatility Trading Using Recurrent Neural Networks," *IEEE Trans. Neural Netw,* vol. 12, no. 4, pp. 865-874, Jul. 2001.
[15] F. Wang, Z. Zhao, X. Li, F. Yu, and H. Zhang, "Stock volatility prediction using multi-kernel learning based extreme learning machine," in *Proc. International Joint Conference on Neural Networks (IJCNN)*, Beijing, China, July 2014, pp. 3078-3085.
[16] S. Colianni, S. Rosales, and M. Signorotti, "Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment Analysis," *CS229 Project*, http://cs229.stanford.edu/proj2015/029_report.pdf, [(Accessed at Mar. 20, 2019), 2015].
[17] A. H. Dyhrberg, "Bitcoin, gold and the dollar – A GARCH volatility analysis," *Finance Research Letters,* vol. 16, pp. 85-92, Feb. 2016.
[18] E. Stenqvist, and J. Lönnö, "Predicting Bitcoin price fluctuation with Twitter sentiment analysis," https://kth.diva-portal.org/smash/get/diva2:1110776/FULLTEXT01.pdf, [(Accessed at Mar. 20, 2019), 2017].
[19] S. Hochreiter, and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation,* vol. 9, no. 8, pp. 1735-1780, Nov. 1997.
[20] D. Park, Y. Hoshi, and C. C. Kemp, "A Multimodal Anomaly Detector for Robot-Assisted Feeding Using an LSTM-Based Variational Autoencoder," in *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, QLD, Australia, Spt. 2018, pp. 1544-1551.
[21] X. B. Peng, M. Andrychowicz, W. Zaremba, and P. Abbeel, "Sim-to-Real Transfer of Robotic Control with Dynamics Randomization," *IEEE Trans. Robot. Autom.,* vol. 3, no. 3, pp. 1544 - 1551, Feb. 2018.
[22] "Yahoo Finance BTC-USD," (Accessed at Mar. 20, 2019); https://finance.yahoo.com/quote/BTC-USD?p=BTC-USD&.tsrc=fin-srch.
[23] "Exchange Volume in 24 Hours," (Accessed at Mar. 20, 2019); https://coinmarketcap.com/exchanges/volume/24-hour/.
[24] A. M. Antonopoulos, *Mastering Bitcoin: unlocking digital cryptocurrencies*: " O'Reilly Media, Inc.", 2014.
[25] "Google Trend," (Accessed at Mar. 20, 2019); https://trends.google.com/trends/?geo=US.
[26] U. Traders, "Top 50 Blockchain and Crypto Twitter Accounts to Follow," https://medium.com/@Uttoken.io/top-50-blockchain-and-crypto-twitter-accounts-to-follow-3a343948d176, [(Accessed at Mar. 20, 2019), 2018].
[27] "TensorFlow," (Accessed at Mar. 20, 2019); https://www.tensorflow.org/install.
[28] "Keras 2.2.4," https://github.com/keras-team/keras/releases.
[29] "U.S. DEPARTMENT OF THE TREASURY," (Accessed at Mar. 20, 2019); https://www.treasury.gov/resource-center/data-chart-center/interest-rates/pages/textview.aspx?data=yield.