Bitcoin's Realized Volatility Forecasting: A Comparison between GARCH and LSTM

He Ma[a,*,#], Xinyi Lin[b,#]

School of Mathematics and Physics, Xi'an-Jiaotong Liverpool University, Suzhou, China

[a]hema.emma22@gmail.com, [b]thea.lin41@gmail.com

[#]These authors contributed equally.

*Corresponding author

Keywords: Bitcoin's volatility, GARCH, GJR-GARCH, Bi-LSTM, Multivariate Bi-LSTM

Abstract: This study compares GARCH and LSTM models, assessing their pros and cons in economic and machine learning. It aims to enhance understanding of their relationship and applicability in various contexts. This article reviews the existing literature on Bitcoin and forecasting models, explains the dataset and methodology used for computing volatility, analyzes GARCH, GJR-GARCH, LSTM, and Bi-LSTM models, introduces three new input dimensions for the Bi-LSTM model, presents predicted outcomes, evaluates them using RMSPE and RMSE, and draws a robust conclusion with insightful findings and future research directions. The investigation reveals that the LSTM model has more outliers but effectively captures Bitcoin's volatility trends, indicating its potential for forecasting complex patterns in the market.

1. Introduction

Ever since its introduction in 2009, Bitcoin has revolutionized the global financial landscape, making its mark as a decentralized cryptocurrency. It has gained recognition as an asset class by a multitude of asset managers, large investment banks, and hedge funds. In addition, some industries are starting to take an interest in the blockchain technology behind Bitcoin and are starting to issue their own cryptocurrencies. For example, several cryptocurrencies have been issued in agribusiness, e.g., Blocery, Carboncoin. As cryptocurrencies in these industries is in an emerging stage, studying the bitcoin market can also help its development in another way (Shen, Wan & Leatham, 2021) [1]

Bitcoin has always been known for its high risk. Looking back at the price data from its launch, Bitcoin's price fluctuation is very evident and has become even more pronounced in recent years, for instance, it plunged nearly 60% in 2022. Therefore, investors need to know enough about its volatility before making investment decisions. Among a series of volatility calculation methods, Bitcoin's return volatility is a simple and efficient measurement, which is quite useful in asset allocation, pricing and risk management. For instance, Bitcoin's return volatility is an important factor when evaluate Value at Risk (VaR), which is developed by JP Morgan's analysts to visualize and measure risk [2].

To forecast volatility, economists usually use traditional economic models, among which GARCHs are widely used. Since it was first proposed in 1986, the GARCH model has been upgraded for more accurate forecasting ability. However, with the continuous advancement of machine learning models, many researchers like Athey (2019) believe that machine learning models will have a huge impact on the financial industry [3]. Compared to traditional economic models that focus on economic theories and principles, machine learning models focus more on data itself. Therefore, the different operating principles allow machine learning models to bring new possibilities to Bitcoin volatility prediction. In particular, RNN models are widely mentioned for their better time series processing capability. Shen (2021) [1] compared the forecasting performance of the GARCH and the RNN on Bitcoin volatility and found that the RNN model didn't perform as well as people might think. In contrast, the GARCH model showed better ability in analyzing extreme market management.

  DOI: 10.25236/icceme.2023.018

However, RNN models are tend to lose information when making certain transformation and LSTM models proposed in 1997 can better detect long-term dependencies of data. Although LSTM has already been used in Bitcoin market, its application is mostly focused on the price prediction. Therefore, this study will compare GARCH and LSTM models to further elaborate the advantages and disadvantages of economic models versus machine learning models.

In the later section, the literature review of Bitcoin and forecasting models is presented. In the third section, details of dataset and calculation method of realized volatility are firstly explained and then it moves on to the methodology of GARCH and GJR-GARCH models. Later, LSTM and Bi-LSTM are proposed. To improve the performance of Bi-LSTM, three new dimensions of input are extracted. The next section shows the predicted results of all models and the root mean squared percentage error (RMSPE) and root mean square error (RMSE) are calculated to evaluate their performance. Finally, conclusion is made in the last section.

2. Literature Review

All literature investigated in this research can be categorized into three parts. The first part focuses on the Bitcoin subject. The second part includes studies about the model building of GARCHs and LSTMs.

In the first section of the literature review, the background of Bitcoin is introduced, according to Wang and Zhou (2022) [4], with CME Group's introduction of Bitcoin futures, options, and micro futures, cryptocurrencies have become a crucial investment for financial institutions. Bitcoin, with the largest market capitalization, holds great significance for portfolio allocation and risk management. Understanding its price dynamics is essential for investors. As mentioned by Tang et al. (2023) [5], due to its rapid fluctuations, Bitcoin has gained significant attention as a benchmark for the digital currency market. However, its high volatility triggers uncertainty within the financial market.

In the second section, the studies of different types of GARCH and LSTM models are included.

For GARCHs, as mentioned by Zahid et al. (2022) [6], the GARCH model is popular for capturing volatility, but it overlooks time-dependent asymmetry. To address this, extensions to the GARCH model have been proposed, considering both shock magnitude and direction. According to Francq (2019) [7], GARCH models focus on conditional variance, which is the variance conditioned on past information. The classical GARCH models express conditional variance as a linear function of the past squared values of the series. This specification effectively captures the essential stylized facts observed in financial series. Kochling et al. (2019) [8] applied different volatility proxies and loss functions to analyse the quality of volatility forecasting of GARCH-type models and finally got a better-performed GARCH model.

For LSTMs, Kazeminia et al. (2023) [9] strived to build a hybrid 2D-CNN LSTM model for Bitcoin price prediction. In this study, the proposed model is more efficient to predict Bitcoin price than CNN, basic LSTM and GRU and is a good supporter for real-time forecasting. In addition, LSTM's priority in Bitcoin price prediction is also approved by Li (2022) [10], who built a multi-features LSTM and states that multidimensional input can improve the performance of LSTM. This is consistent with Chen et al. (2022) [11], who point that multi-dimension features can make the learning process more relevant to the reality. Prakash et al. (2023) [12] use Prophet and another five types of LSTM models to predict COVID-19 Pandemic, for instance, Bi-LSTM, CNN-LSTM and GRU-LSTM. Although the topic is not that relevant, it provides significant reference in terms of LSTM model's type choosing, building and improvement.

3. Materials and Methodology

3.1. Dataset Details

The historical dataset of Bitcoin Open, Close, High, and Low prices is obtained from finance.Yahoo.com and it ranges from 1 January 2020 to 7 June 2023. In the research, bitcoin's daily return will be calculated. For practical purposes, the log return is used to reduce non-stationary data. The formula is as follows:

$$r_{t,t+1} = log(Pt + i/Pt)$$

Before constructing GARCH models, the Augmented Dickey-Fuller (ADF) T is used to conduct the data stationary test. The ADF test is an extension of the original Dickey-Fuller test, which is only suitable for first-order tests. An autoregressive process can be expressed as follows:

$$y_t = by_{t-1} + \alpha + \epsilon_t$$

If the coefficient 'b' of the lag term is 1, it becomes a unit root. When the unit root exists, the series is not smooth and is a random walk process. When the unit root does not exist, the series is smooth and passes the ADF test.

3.2. Realized Volatility

Bitcoin Volatility is a measure that reflects Bitcoin's price fluctuates over a certain period of time. There are several different proxies for realized volatility calculation and the most frequently used one is calculating the standard deviation of price return. The formulas are as follows:

$$\sigma_{daily} = \sqrt{\sum {r_{t,t+1}}^2} * \sqrt{\frac{1}{i-1}}$$

where σ_{daily} is the daily volatility, $r_{t,t+1}$ is the daily return, *i* represents the fixed interval window.

According to the formulas, the daily volatility will be influenced by the length of the interval. However, a fixed and proper interval is needed in this research. To determine a suitable parameter, five intervals (7, 30, 60, 180, 365) are brought into comparison. It can be seen from Figure 1 that volatilities with 7-Day intervals are too noisy to extract information effectively, while those with longer interval windows are too smooth and are largely close to the mean. Therefore, a 30-Day interval Realized Volatility seems to be the best choice, as it also prevents wasting the first few data.

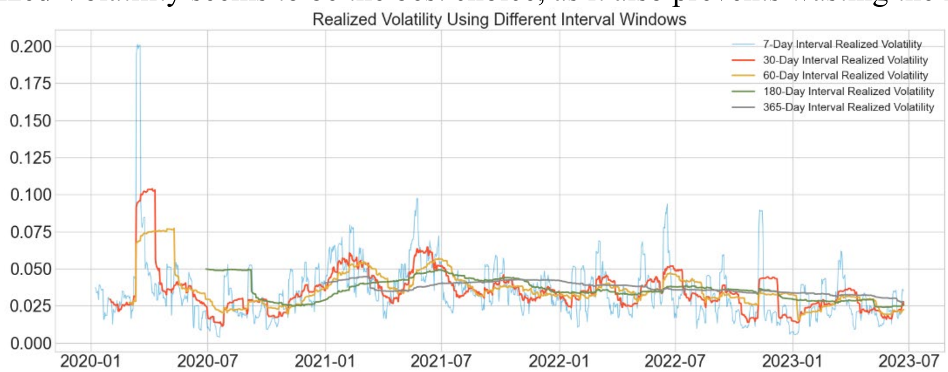

Figure 1 Realized volatility using different interval windows.

3.3. GARCH Model

The GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model is an extension of the ARCH (Autoregressive Conditional Heteroskedasticity) model. The GARCH model aims to predict future volatility by utilizing past volatility data. It assumes that the volatility in financial time series is not constant and exhibits autoregressive properties. The GARCH model consists of two main components. First, it incorporates a conditional mean model, which is used to predict the mean of the financial time series. The second is the conditional variance model, which is used to predict the volatility of the financial time series. The general form of the GARCH model can be expressed as follows:

$$\alpha_t^2 = \omega + \sum_{i}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{1}^{p} \beta_i \sigma_{t-i}^2$$

where σ_t^2 is variance at time step t, ϵ_{t-i}^2 is the model residuals at time step t-1 and α plus β is equal to 1. Moreover, w is the long-term variance.

To determine the appropriate orders (p and q) for the GARCH model, Returns Autocorrelation and Returns Partial Autocorrelation are utilized. Once the orders are determined, the GARCH model equation mentioned above is employed to forecast the volatility of the financial time series. It is worth noting that the volatility of the second day is closely related to that of the first day. Empirical studies have shown that the basic GARCH model assumes symmetric positive and negative news effects on volatility. However, financial time series data often deviate from a normal distribution, exhibiting a higher likelihood of extreme values that deviate from the mean. In other words, the effects tend to be asymmetric, with negative effects being greater than positive ones.

To address this issue, the GJR-GARCH (short for Glosten-Jagannathan-Runkle GARCH) considers the asymmetry of shock responses, which GARCH lacks. The GJR-GARCH model can be expressed as follows:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} (\alpha_i + \gamma_i N_{t-i}) \alpha_{t-i}^2 + \sum_{i=1}^{s} \beta_j \sigma_{t-j}^2$$

where N_{t-i} is a schematic function denoting a_{t-i}<0 and $\alpha_i, \gamma_i, \beta_j$ are non-negative parameters.

3.4. LSTM Model

Long Short-Term Memory Network (LSTM) proposed in 1977, is an improved version of RNN that can identify long-term dependencies in the data. A basic LSTM consists of four parts: a cell, an input gate, a forget gate, and an output gate. The cell is responsible for memorizing information and the other three gates are charged with managing the flow of information in and out of the cell. Forget gates help to filter useless information from the previous step. Input gates can select new information for the current step and output gates will pick information that is suitable for outputting to the next step. Therefore, LSTMs are good at learning to judge the importance of data and storing valuable information in the long-term stage and are the ideal neural network model for time series forecasting.

Bidirectional LSTM: Bi-LSTM is an extension of the LSTM cell. The model contains and trains two LSTM models at the same time, with one of them processing input forward and the other one running in reverse. Due to the extra backward layer, it can provide more additional information to the networks and result in more accurate and relevant results.

Multivariate bidirectional LSTM: In order to provide adequate data and yield better results, the dimension of inputs should be enlarged. Therefore, instead of using daily volatility as the only kind of input, the open/high/low/close prices and volume are included as relevant data. However, to avoid high correlation, three features are extracted:

$$Open\text{-}Close\ Spread = \frac{(Close - Open)}{Close} * 100\%$$

$$High\text{-}Low\ Spread = ln\frac{(High - Low)}{Close} * 100\%$$

$$Daily\text{-}Volume = ln(Volume)$$

In addition, dropout layers are added to reduce the possibility of overfitting and increase its performance on the Validation Set.

4. Experiment Analysis

In this section, models' performance will first be compared separately within the GARCHs group and the LSTMs group, and then a comparison between the two types of models, GARCH and LSTM, will be conducted.

To evaluate the models' performance in Validation Set, the root mean squared percentage error (RMSPE) and root mean square error (RMSE) are evaluated. However, RMSPE is prioritized because the value of volatility is highly influenced by the interval window selection and the form of percentage can minimize the effect caused by value changes. In addition, RMSPE is more sensitive to outliers compared with regular MAPE.

4.1. Performance of GARCHs

Table 1 shows two GARCH models are built in the research. Model I is GARCH(1,1) and Model II represents GJR-GARCH(1,1,1). Their performance results are showed as below:

Table 1 Performance of two GARCH models (RMSPE and RMSE).

	Model I	Model II
RMSPE	0.703420	0.495009
RMSE	0.100360	0.060441

Based on the above table, it can be seen that forecasts of Model Ⅱ have shown significant improvement, with a much smaller RMSPE. As the detailed prediction results are showed in Figure 2, it can be seen that forecasts of Model Ⅱ is closer than the target value. Although between February and June 2023, forecasts of Model Ⅰ and Model Ⅱ show nearly identical fluctuation and clustering effects, but apparently Model Ⅰ underestimated volatility, which may be the main cause of its high RMSPE.

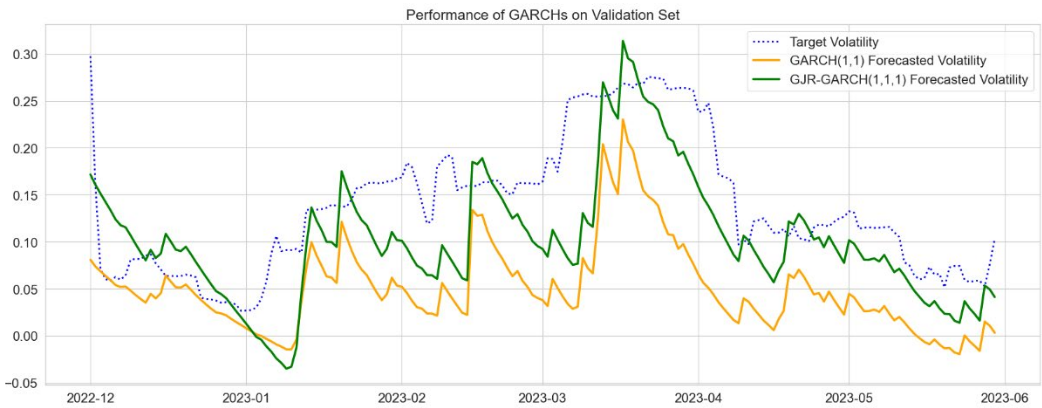

Figure 2 Performance of GARCHs on Validation Set.

4.2. Performance of LSTMs

Table 2 shows two LSTM models are proposed in the research. To make the comparison effect more obvious, a basic LSTM model is created. Model 1 is the basic LSTM with only 1 hidden layer of 20 units. Model 2 is the 2-layered Bi-LSTM with 32 and 16 units in layers respectively. The multivariate Bi-LSTM (Model 3) has three hidden layers (64, 32, 16) and a dropout layer with 0.1. Their performance results are showed as below:

Table 2 Performance of three models (RMSPE and RMSE).

	Model 1	Model 2	Model 3
RMSPE	1.018545	0.786468	0.736277
RMSE	0.067883	0.054780	0.062278

Comparing the RMSPE, Model 3 performs best, which might imply that the multivariate Bi-LSTM model generates fewer outliers than the other two models. However, the RMSPEs of Model 2 and Model 3 are very close and Model 2 outperforms both Model 1 and Model 3 in RMSE. To further analyse the two models, their prediction results are showed below.

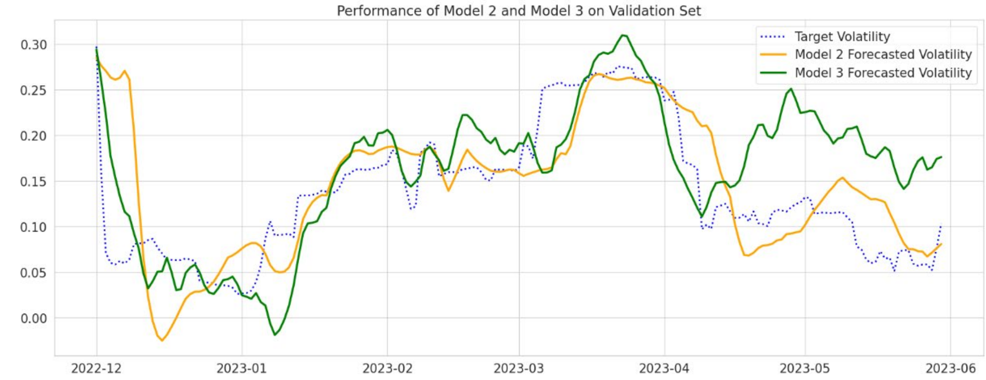

Figure 3 Performance of Model 2 and Model 3 on Validation Set.

According to the figure, Model 3 seems to be more sensitive to fluctuations and can better capture the volatility trends than Model 2. It might be one explanation for the lower RMSPE of Model 2 since it tends to be lagging behind compared to the desired target and hence generate more outliers. However, Model 3 produces a large deviation in the final stage (from 10 April 2023 to 30 May 2023), with the predicted value much larger than the target one, which may be the reason why its RMSPE is not outstanding and RMSE is outperformed by Model 2.

4.3. Comparison between GARCH and LSTM

From the perspective of RMSPE, all two GARCH models overperform LSTM models, implying that LSTMs might generate more outliers than GARCHs. However, LSTMs seem to better capture the varying trend of bitcoin volatility according to Figure 2 and Figure 3. Although during the final period of time, the multivariate Bi-LSTM largely overvalues volatility, its prediction of trends still matches the target. In contrast, results of GARCHs show a stronger effect of clustering, commonly called as “volatility clustering”. The phenomenon means that high volatility tends to be followed by large turbulence and low volatility is usually accompanied by a peace period. It is very common in financial time series, but it may not fit the target volatility between 1 December 2022 and 30 May 2023 in Validation Set, since Bitcoin is in a relatively low and stable period compared to 2021 and the first half of 2022.

5. Conclusion

Bitcoin, the world's most successful virtual currency, has a market capitalization of more than $870 billion. However, high returns always come with high risk. Bitcoin prices fluctuate much more severely than the vast majority of financial assets. Therefore, forecasting volatility is very important for investors when making investment decision. In this research, both GARCHs and LSTMs are built to forecast the bitcoin return volatility.

The LSTM model is considered to have a more prominent and excellent performance. Early academic reports on bitcoin price prediction have demonstrated this. However, by comparing the forecasting performance on Validation Set, LSTMs seems to produce more outliers than GARCHs and the accuracy of forecasts are also outperformed. In addition, to better improve its accuracy, Bi-LSTM model with multivariate inputs is built, but from the results, its accuracy did not improve greatly. But according to the line graph, LSTMs seems to be more sensitive to fluctuations and can captures the trend more accurately.

The study makes a further comparison between economic models and machine models. Different types of GARCHs are built to catch their best results, while LSTMs are improved to make the learning results more accurate. The results of this research are consistent with the previous study that machine learning (RNN) will generate more outliers and can corresponds more quickly to the volatility vibration [3]. However, this research improves the structure of both the GARCH and the LSTM models, and the overall forecasting accuracies are improved, since RNN model largely underestimate the volatility in the former research.

References

[1] Ze Shen, Qing Wan, & David J. Leatham. (2021). Bitcoin Return Volatility Forecasting: A Comparative Study between GARCH and RNN. *Journal of Risk and Financial Management*, *14*(337), 337. https://doi.org/10.3390/jrfm14070337

[2] Uyar, U., & Kahraman, I. K. (2019). The risk analysis of Bitcoin and major currencies: value at risk approach. *Journal of Money Laundering Control*, *22*(1), 38–52. https://doi.org/10.1108/JMLC-01-2018-0005

[3] Athey, S. (2019). 21. The Impact of Machine Learning on Economics. In A. Agrawal, J. Gans & A. Goldfarb (Ed.), *The Economics of Artificial Intelligence: An Agenda* (pp. 507-552). Chicago: University of Chicago Press. https://doi.org/10.7208/9780226613475-023

[4] H. Wang & X. Zhou. (2022), "Less is More: Bitcoin Volatility Forecast Using Feature Selection and Deep Learning Models," *2022 IEEE 20th International Conference on Industrial Informatics (INDIN)*, (pp. 681-688). Perth, Australia. doi: 10.1109/INDIN51773.2022.9976100.

[5] Tang, X., Song, Y., Jiao, X., & Sun, Y. (2023). On Forecasting Realized Volatility for Bitcoin Based on Deep Learning PSO–GRU Model. *Computational Economics*, 1–23. https://doi-org-s.elink.xjtlu.edu.cn:443/10.1007/s10614-023-10392-5

[6] Zahid, M., Iqbal, F., & Koutmos, D. (2022). Forecasting Bitcoin Volatility Using Hybrid GARCH Models with Machine Learning. *Risks*, *10*(12), 237. https://doi-org-s.elink.xjtlu.edu. cn:443/10. 3390/risks10120237

[7] Francq, C. (2019). *GARCH models : structure, statistical inference and financial applications* (J.-M. Zakoian (Ed.); Second edition.). John Wiley & Sons

[8] Köchling, G., Schmidtke, P., & Posch, P. N. (2020). Volatility forecasting accuracy for Bitcoin. *Economics Letters*, *191*. https://doi-org-s.elink.xjtlu.edu.cn:443/10.1016/j.econlet.2019.108836

[9] Kazeminia, S., Sajedi, H., & Arjmand, M. (2023). Real-Time Bitcoin Price Prediction Using Hybrid 2D-CNN LSTM Model. *2023 9th International Conference on Web Research (ICWR), Web Research (ICWR), 2023 9th International Conference On*, 173–178. https://doi.org/10.1109/ICWR57742.2023.10139275

[10] Li, T. (2022). Prediction of Bitcoin Price Based on LSTM. *2022 International Conference on Machine Learning and Intelligent Systems Engineering (MLISE), Machine Learning and Intelligent Systems Engineering (MLISE), 2022 International Conference on, MLISE*, 19–23. https://doi.org/10.1109/MLISE57402.2022.00012

[11] Chen, Z., Li, C., & Sun, W. (2020). Bitcoin price prediction using machine learning: An approach to sample dimension *engineering. Journal of Computational and Applied Mathematics, 365.* https://doi.org/10.1016/j.cam.2019.112395

[12] Prakash, S., Jalal, A. S., & Pathak, P. (2023). Forecasting COVID-19 Pandemic using Prophet, LSTM, hybrid GRU-LSTM, CNN-LSTM, Bi-LSTM and Stacked-LSTM for India. *2023 6th International Conference on Information Systems and Computer Networks (ISCON), Information Systems and Computer Networks (ISCON), 2023 6th International Conference On*, 1–6. https://doi.org/10.1109/ISCON57294.2023.10112065