Bitcoin Return Volatility Forecasting: A Comparative Study of GARCH Model and Machine Learning Model

Ze Shen[1], Qing Wan[2], David J. Leatham[3]
1. Department of Agricultural Economics, Texas A&M University
sz910201@tamu.edu
2. Department of Computer Science & Engineering, Texas A&M University
3. Department of Agricultural Economics, Texas A&M University

Selected Paper prepared for presentation at the 2019 Agricultural & Applied Economics Association

Annual Meeting, Atlanta, GA, July 21-23

Abstract:

One of the well-known features of bitcoin is its extreme volatility. The modeling and forecasting of bitcoin volatility is crucial for bitcoin investors' decision making analysis and risk management. All the previous studies of bitcoin volatility were founded on economic models. However, research on bitcoin volatility forecasting using machine learning algorithms is still void. In this article, both conventional economic models and machine learning model are used to forecast the volatility of bitcoin return. The objective of this study is to compare their out-of-sample performance. The results demonstrate recurrent neural network method outperforms the economic GARCH model and simple moving average model. Then it provides more motivation for the economic researchers to apply machine learning methods to the financial and economics world.

Keyword: bitcoin, GARCH, machine learning, recurrent neural network, volatility

JEL codes: G000, G170

Since Satoshi Nakamoto proposed the first cryptocurrency in 2009, the cryptocurrency market has received much attention. Bitcoin is the most successful and popular one in the market, which accounts for over fifty percent of the current whole cryptocurrency market capitalization.[1] The bitcoin enthusiasm is due to its innovational features of decentralization, anonymity and zero transaction cost. Researchers analysis of bitcoin has recently received growing interests. David Yermack (2015) studied the features and functions of bitcoin, and concludes that bitcoin appears to be more like a speculative investment than a real currency due to its high volatility. If people look at the bitcoin price history from 2009 till now, its violent fluctuations will be discovered. As a financial asset, bitcoin is famous for its extreme volatility. The modeling and forecasting of bitcoin volatility is crucial for bitcoin investors' decision making analysis and risk management.

Earlier studies mainly explored bitcoin volatility by using GARCH family models. Bouoiyour and Selmi (2015, 2016) compared different GARCH type model on sub-period bitcoin volatility, and Paraskevi (2017) compared the GARCH family models over the whole period. Mehmet el at. (2017) found the bitcoin trading volume fails to predict bitcoin volatility by studying their causal relationship.

These early studies of bitcoin volatility were founded on economic models. However, research on bitcoin volatility forecasting using machine learning algorithms is still void. Susan Athey (2018) pointed out that the machine learning would have a dramatic impact on the field of economics in the near future. Unlike the economic models, where researcher picks a specific model based on economic principles and estimates the parameters, machine learning algorithm is usually a data driven modeling focused on the selection process. Thus a model of machine learning algorithm is

[1] From coinmarket.com, the total cryptocurrency market capitalization was around $1,800 billion and bitcoin market capitalization was over $97 billion on April 2019.

not fixed or predetermined but will be refined during a training process. Applying machine-learning methods to solve for economic issues can potentially make a difference in the economic and financial field.

In this article, both a conventional economic model and a machine learning model are used to forecast the volatility of bitcoin return, and their forecasting performance are evaluated. The aim of this article is to compare their performance, and to discover if machine learning can improve economic time series forecasting. The booming development of machine learning techniques in time series forecasting encourages people to apply it in financial market. Moreover, the success of machine learning on stock market prediction leads us to believe that it may work well for cryptocurrency price forecasting. In addition, the empirical studies show that the machine learning method is more efficient than ARIMA model in bitcoin price prediction. Sean, Jason and Simon (2018) compared the forecasting performance of recurrent neural network (RNN), long short term memory (LSTM) network and ARIMA on bitcoin price, and reported that the machine learning models outperformed ARIMA. Laura A. et at (2018) examined the forecasting performance on cryptocurrency portfolio, and reported the same conclusion that machine learning methods overwhelms the standard benchmark simple moving average. It makes sense for machine learning method to be superior to traditional economic model (such as simple moving average and ARIMA), because machine learning model is proposed in a more general scope that takes both linear and nonlinear features into consideration. It also preserves more temporal information of a time series during training.

The machine learning methods are more advanced than some traditional economic models in time series forecasting, both theoretically and empirically. However, this assertion needs to be cautious. First of all, the economic models involve economic intuition while machine learning mainly deals with data. In an economic world, the economic intuition is the key to economic analysis. In contrast, the machine learning captures information only from data. However, the information contained in data is limited in analyzing economic issues. Secondly, the performance of machine learning depends on large amount of data. The performance is dramatically improved as the data amount getting larger. However, in this article, the bitcoin market history is quite short and the most frequency data available is the daily data. Finally, machine learning is sensitive to the fluctuations. Compared to other approaches, machine learning is more efficient in identifying time series trends and patterns. However, this leads to a problem that a shock or abnormal perturbation will be treated more seriously. But in the real world, there are many factors that affecting the market reaction to the shock or abnormal perturbation, the fluctuation sensitivity might cause overreaction problem in the forecasting, especially for the volatility analysis.

The objective of this study is to compare the forecasting performance between traditional economic models and machine learning method. The contribution of this article is to investigate whether the machine learning method is more advanced in bitcoin volatility forecasting and how advanced it is going to be. First, the economic model is presented. The article starts with the naive model, simple moving average model as a benchmark, and then move to a more complex but conventionally applied model, generalized autoregressive conditional heteroscedasticity (GARCH) model, to forecast bitcoin return volatility. Then a machine learning model based on Recurrent Neural Network (RNN) is proposed. The next step is to evaluate the out-of-sample performance of the three models. The root mean squared error (RMSE) and mean absolute error (MAE) are used to evaluate and compare their performances. Since the true conditional volatility of bitcoin return is unobservable, Garman-Klass volatility (Garman and Klass, 1980) is used as a proxy for the realized volatility.

Data

In this study, the bitcoin return time series is used rather than the raw bitcoin price data. The bitcoin daily return is defined as the difference of the natural logarithm of the daily bitcoin closing price. Bitcoin daily opening, high, low and closing price are used to estimate the realized bitcoin volatility. All the data are available in website: CoinMarketCap.com. The data ranges from April 30, 2013 to November 20, 2018, 2031 observations totally. Figure 1 and Figure 2 illustrates the bitcoin daily return and bitcoin daily squared return respectively. Table 1 shows the descriptive statistics of the bitcoin daily return.

Table 1. Summary statistics for bitcoin daily returns in the sample period

Sample size	2031
Mean	0.000733
Variance	0.000361
Std.	0.019004
Skewness	-0.195558
Kurtosis	8.017161

Figure 1. BTC daily return

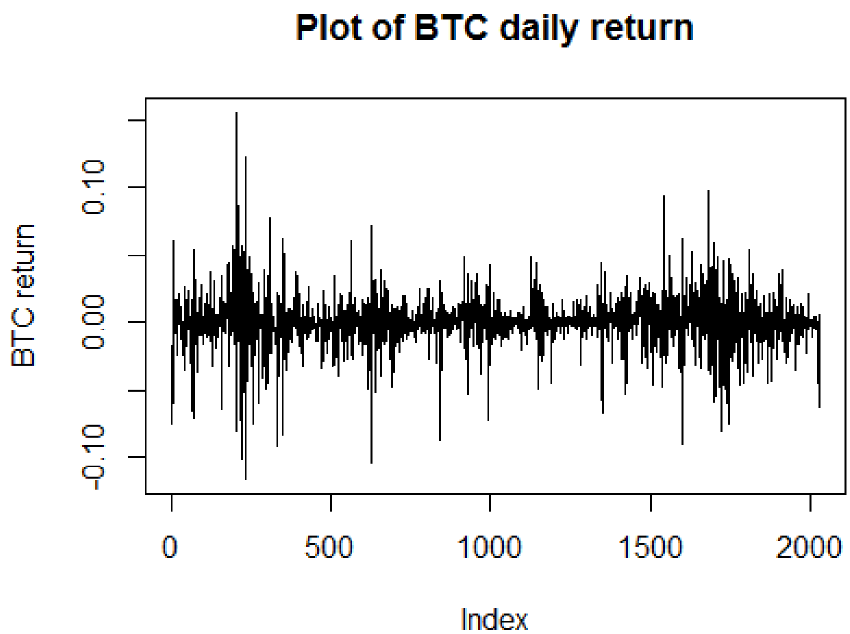
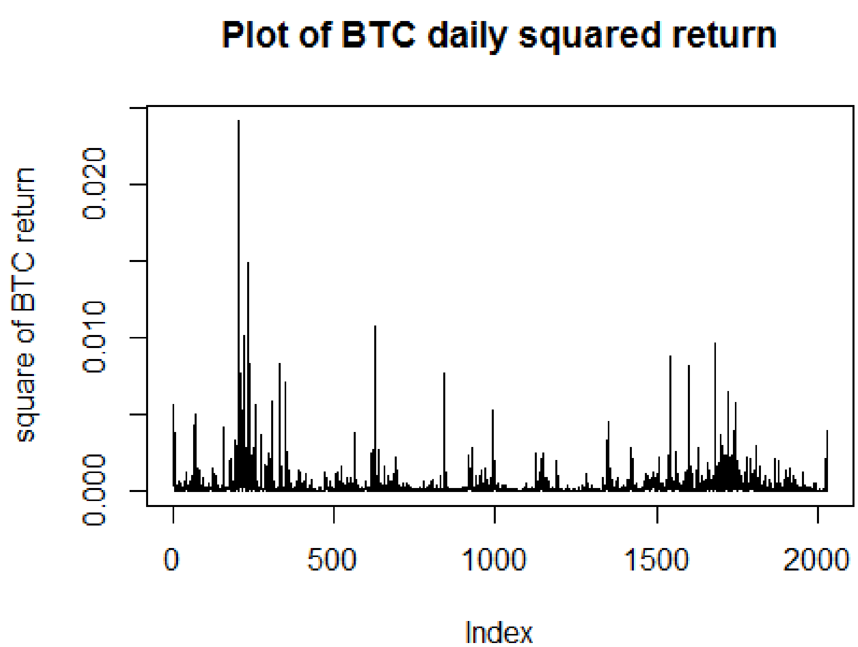


Figure 2. BTC daily squared return



Before going further to the economic modeling, the stationary of time series must be checked. The augmented Dickey-Fuller- test (ADF) and Phillips-Perron (PP) unit root test are used to check for the stationary of bitcoin daily return series, and table 2 indicates that the financial time series is stationary.

Table 2. Unite Root Tests.

	Without Trend		With Trend	
	ADF	PP	ADF	PP
BTC daily return	-44.52	-44.74	-44.51	-44.73
Critical values (1%)	-3.43	-3.43	-3.96	-3.96

Methodology

In this section, the economic methodology is discussed first, and then the recurrent neural network model, which is a machine learning methodology will be presented. Engle in 1982 proposed the autoregressive conditional heteroscedasticity models (ARCH), which assumes that the volatility of asset returns is time varying instead of a constant. Bollerslev (1986) generalized the ARCH model and developed a more commonly used GARCH model. In this study, the GARCH model is applied as the economic method.

Economic methodology

First look at figure 1, as it shows notable fluctuations in bitcoin daily return. It is also found that large changes follow large turbulence and small changes follow calm periods. This phenomena in time series asset return is known as “volatility clustering”. The plot of bitcoin daily squared return in figure 2 provides more evidence that changes tend to be cluster together.

Figure 3 shows the autocorrelation function and partial autocorrelation function of bitcoin daily squared return. Table 4 shows the results of Ljung-Box Q-test for the bitcoin daily squared return. Figure 3 and table 4 indicate that the bitcoin daily squared return is serially correlated, which suggests the existence of the conditional heteroscedasticity in bitcoin price volatility. Thus, the economic model needs to capture the feature of heteroscedasticity.

Figure 3. ACF and PACF of BTC daily squared return

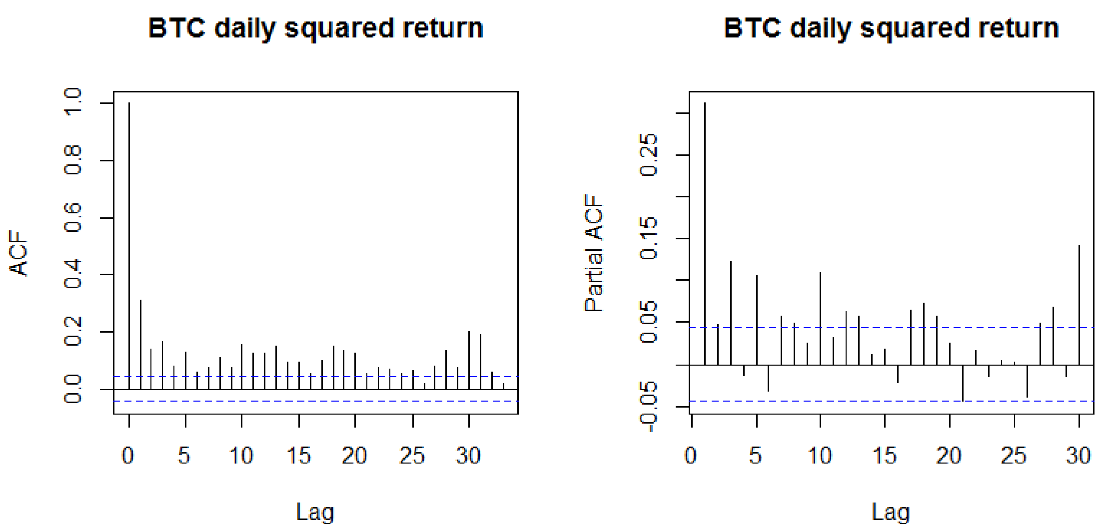


Table 3. Ljung-Box Q-test for BTC daily return

No. of lags	lag 10	lag 15	lag 20
P-value	0.006623***	0.005274***	0.00005***

Note: *** denotes for the significance at 1% level

The basic structure of the economic model is as follows:

$$r_t = \mu_t + Z_t \quad (1)$$

$$\mu_t = E(r_t | F_{t-1}) \ (2)$$

$$h_t^2 = Var(r_t | F_{t-1}) = E[(r_t - \mu_t)^2 | F_{t-1}] = E(Z_t^2 | F_{t-1}) \ (3)$$

where $r_t = \frac{\log P_t}{\log P_{t-1}}$, μ_t is the conditional mean and h_t^2 is the conditional variance, $\digamma_{t-1}$ denotes for the past information.

Conditional mean

ARMA(p, q) process is applied to model the conditional mean:

$$\mu_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + \sum_{j=1}^{q} \theta_j Z_{t-j} \quad (4)$$

With the autoregressive order p and moving average order q.

After applying the ARMA(p, q) process, the estimated parameters and the residuals are obtained. As it is discussed above, the bitcoin daily return exhibits volatility clustering, which indicates the conditional heteroscedasticity volatility. The ARCH effects of the residuals are tested. If there is ARCH effect in the residuals, the conditional variance models will be specified in the next section.

Conditional variance

Given the conditional mean model and using (3), the residuals Z_t, $Z_t = r_t - \mu_t$ are obtained. Then the condition variance models are able to be built. Two different models are presented in the following section. It start with the naïve model, a simple moving average model, and then move to a more complex but conventionally applied model, GARCH model, to forecast bitcoin return volatility.

Simple moving average

Even though simple moving average is the simplest model for volatility forecasting, it models the time varying variance and captures the past information and historical variance. Although the simple moving average model incorporates neither conditional mean nor conditional variance in the sense of GARCH, it is presented here as a benchmark to evaluate the performance of the other models.

The simple moving average model is presented as:

$$\sigma_{k+1}^2(n) = \frac{1}{n}\sum_{i=0}^{n-1} r_{k-i}^2 \quad (5)$$

Where k is the forecast origin, and r_t^2 is the bitcoin daily squared return. We set $n = 10$ in this paper.

GARCH model

The generalized autoregressive conditional heteroscedasticity model (GARCH) was developed by Bollerslev in 1986. Both ARCH process and GARCH process model the variations of a financial assets' volatility, and the GARCH process allows the conditional variance to be an ARMA process. The GARCH (m, n) process is as follows:

$$Z_t = h_t \varepsilon_t, \{\varepsilon_t\} \sim IID(0,1) \quad (6)$$

$$h_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i Z_{t-i}^2 + \sum_{j=1}^{n} \beta_j h_{t-1}^2 \quad (7)$$

$\{Z_t\}$ is the residual series of the best fitting ARMA(p, q) model, thus the conditional variance of the residual series essentially acts like an ARMA process. It is then expected that the standardized squared residuals obtained from best fitted ARMA-GARCH model should not be autocorrelated and there should not remain any ARCH effects. The ARCH LM test is used to check whether this is true or not.

Recurrent neural network

Our sequencing model for predicting bitcoin return volatility is built on the concept of Recurrent Neural Network (RNN). RNN deals well with sequence problems, which thanks to its special architecture that takes the order of data into consideration. Each RNN has a type of memory unit concatenated into multi stages and each of which will turn previous states and current input to activations and pass necessary information forward to the next stage. In this study, a GRU (Gated Recurrent Units) cell is employed for serving as the memory unit. The cost function is redesigned based on a tangent function. This model doesn't build any embedding or probability layer inside that are usual configurations that exist in some engineering task. In addition, by consider some uncertainty of the volatility, the range is equally cut into 250 intervals to convert a real volatility value to a vector with a dimension of 250. This conversion serves as an encoder for a RNN cell's input. A whole architecture of our model is listed as Fig 4. In general, the encoding process will turn a fixed length of sequential data into the same length of vectors for RNN, which is fed into multiple layers of perceptron (MLP). The MLP will decode states from RNN into sequential vectors and transfer them to a predictor for output.

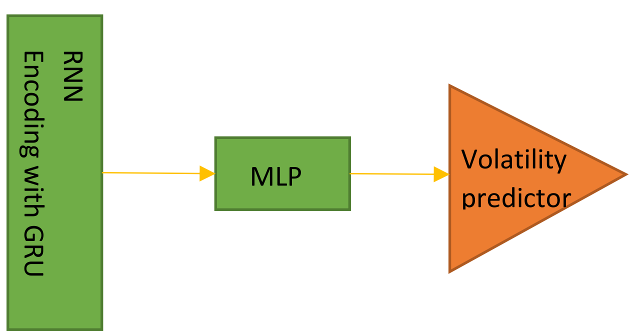


Fig 4

Empirical results

In this section, the forecasting results of the simple moving average model, GARCH model and the recurrent neural network model will be presented. Then, their out-of-sample performance will be evaluated and compared. However, before the evaluation, an appropriate proxy for the realized volatility has to be found.

Forecasting

The sample data is divided into two parts, in sample period from April 30, 2010 to April 30, 2018 (1827 observations) and out of sample period from May 1st, 2018 to November 20, 2018 (204 observations).

In economic GARCH model, the ARMA (p, q) order are selected by AIC and BIC, and the best fitted conditional mean model was found to be ARMA$(2,2)$. Then, the ARCH effects of the residuals are tested, and the result indicates there remains ARCH effect in the residual series. Finally the best fitted ARMA $(2,2)$-GARCH$(1,2)$ model is obtained. Table 4 presents the estimated parameters of ARMA $(2,2)$-GARCH$(1,2)$ model.

Table 4. ARMA (2,2)-GARCH(1,2)

Parameters	Estimated value	t-value	p-value
ϕ_0	0.001	3.734	0.000***
ϕ_1	1.495	133.042	0.000***
ϕ_2	-0.948	-102.038	0.000***
θ_1	-1.513	-121.275	0.000***
θ_2	0.957	222.547	0.000***
α_0	0.000	1.842	0.066*
α_1	0.224	7.720	0.000***
β_1	0.367	2.441	0.015**
β_2	0.408	3.014	0.003***

Note: "***", "**" and "*" denotes for the significance at 1%, 5% and 10% level, respectively.

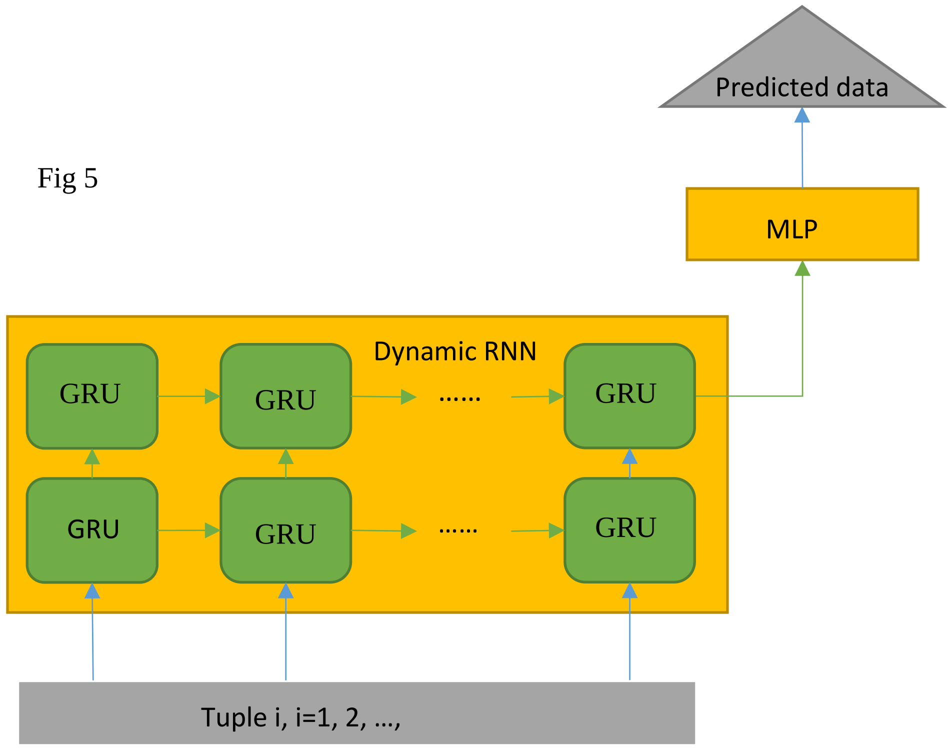
The autocorrelation in the standardized residuals of the fitted ARMA-GARCH model is checked, and the result indicates that there is no remaining ARCH effect in the residuals.

For the recurrent neural network model, 30 days samples of the volatility are used to predict the next 1 day with an out of sample method. For example, the first 30 days of volatility values was used to predict the 31^{st}. The sequential data generated by this process is called as tuple 1; then the 2^{nd} to 31^{st} volatility values are used to predict the 32^{th}, and it is called tuple 2. The total data length was 2031. By rolling this process, 1994 tuples were generated. The out of sample method, the first 1794 (90%) observations were appointed to training and remaining 200 observations were used as a test volume. Also, due to value of volatility being very small, each value was scaled by 10^4.

A more detailed implementation is illustrated in Fig 5. We built two layers of RNN with GRU cell as core. The first layer has 512 units while second shrinks to 256 units. Sequential data were fed in cells on the bottom from left to right. The predicated data were collected on the top from left to right.

Both training and test were taken on GTX 1070 GPU. A SGD (Stochastic Gradient Descend) algorithm that shuffles the whole dataset is used in each iteration; RMSProp gradient update algorithm was chosen as an optimizer; learning rate and batch size was set to 0.0001 and 20, respectively. As it stated before, the model 1000 epochs is trained on the 1794 tuples and the 200 tuples are tested every five epochs.

Fig 5



Volatility proxies

One difficulty of evaluating the forecasting performance is that the true conditional volatility of bitcoin return is unobservable. Thus, a proxy for the realized bitcoin return volatility has to be found. The most common used proxy for the volatility is the bitcoin daily squared return, however, it will lead to a poor out-of-sample performance (Anderson and Bollerslev, 1998). The cumulative squared intra-day returns is a more efficient proxy for volatility (Chou et al., 2010), but it requires high frequency bitcoin prices in one day, which is not available in our case. Thus, the Garman-Klass volatility (Garman and Klass, 1980) is used as the proxy for bitcoin return volatility. This includes the information of daily high, low, opening and closing prices. Garman and Klass (1980)'s estimator in practical is presented as:

$$\hat{\sigma}_{GK}^2 = 0.5[\ln(BTC_{Ht}/BTC_{Lt})]^2 - [2\ln 2 - 1][\ln(BTC_{Ct}/BTC_{Ot})]^2 \quad (8)$$

Where $BTC_{Ht}$and BTC_{Lt} is the highest bitcoin price and lowest bitcoin price at the trading day; BTC_{Ct} and BTC_{Ot} is the closing price and opening price respectively.

Out-of-sample performance

To compare the out-of-sample performance of our three models, the root mean squared error (RMSE) and mean absolute error (MAE) are used to evaluate and rank them. Table 5 exhibits the out-of-sample performance of the simple moving average (benchmark), GARCH model and RNN model. The results of two criterion (RMSE and MAE) are consistent. The RNN model performs best with the lowest RMSE and MAE, and the benchmark SMA performs worst. The RMSE and MAE of the GARCH model are 25.5% and 8% larger than that of RNN model, respectively. The RMSE and MAE of the benchmark simple moving average (SMA) are 0.8% and 1.9% higher than that of GARCH model.

Table 5. Out-of-sample Performance

	RMSE	MAE
SMA	0.00402	0.00165
GARCH	0.00399	0.00162
RNN	0.00318	0.00150

As expected, the RNN model is more efficient in bitcoin return volatility forecasting. The GARCH model improves the forecasting accuracy over the simple moving average, but is overwhelmed by the machine-learning model.

Conclusion

Bitcoin is the most successful and popular cryptocurrency in the market, with around 130 billion daily trading volume as of April 2019. Bitcoin has historically had a larger fluctuations in price than most other financial assets. Thus the analysis of bitcoin return volatility is crucial for investors' decision making and risk management. Both the GARCH model and the recurrent neural network method are used to forecast the bitcoin return volatility.

The machine learning methods in time series forecasting are expected to be superior to the traditional economic models. The earlier empirical studies in stock price forecasting and cryptocurrencies prices forecasting provided evidence this is true. However, it is skeptical of this assertion with the three questions proposed in the beginning. Compare the out-of-sample performance of each model, and the result indicates that compared to the traditional economic model, the machine learning method is more accurate in bitcoin return volatility forecasting, which is consistent with the results of financial price forecasting studies.

One advantage of recurrent neural network is that it is nonlinear model, and the model is learned through past experience. If there is enough data for recurrent neural network to learn, it should outperform the linear models, such as GARCH model. In our study, 1826 in sample observations are sufficient for the model to learn from experience and use them to predict the future.

However, one concern on machine-learning models is that the economic models consider economic backgrounds while machine-learning models only deal with data. This might lead to a problem that machine-learning models are less efficient in forecasting. However, in this article, the RNN model makes more accurate forecasting by using only price data, which suggests that the financial price data also contains many market information. These information are captured by the RNN model and are fully used to improve the forecasting accuracy. The last concern is that the fluctuation sensitivity of machine learning model would cause overreaction problem, which has negative impact on its volatility forecasting performance. Although we are unable to track whether this feature indeed hurt the forecasting accuracy or not, the recurrent neural network model outperforms the GARCH model anyway.

This study proposed an alternative way of volatility analysis. It illuminates the feasibility and potentialities to apply machine learning methods to economic time series forecasting.

Reference

Andersen, T.G., and T. Bollerslev "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts." *International economic review* (1998):885-905.

Athey, S. "The impact of machine learning on economics." *The Economics of Artificial Intelligence: An Agenda.* Anonymous University of Chicago Press, 2018.

Balcilar, M., E. Bouri, R. Gupta, and D. Roubaud "Can volume predict Bitcoin returns and volatility? A quantiles-based approach." *Economic Modelling* 64 (2017):74-81.

Bertolami, R., H. Bunke, S. Fernandez, A. Graves, M. Liwicki, and J. Schmidhuber "A Novel Connectionist System for Improved Unconstrained Handwriting Recognition." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31 (2009):

Bollerslev, T. "Generalized autoregressive conditional heteroskedasticity." *Journal of Econometrics* 31 (1986):307-27.

Bouoiyour, J., and R. Selmi "Bitcoin: A beginning of a new phase." *Economics Bulletin* 36 (2016):1430-40.

Bouoiyour, J., R. Selmi, A.K. Tiwari, and O.R. Olayeni "What drives Bitcoin price." *Economics Bulletin* 36 (2016):843-50.

Chou, R.Y., H. Chou, and N. Liu "Range volatility models and their applications in finance." *Handbook of quantitative finance and risk management.* Anonymous pp. 1273-1281. Springer, 2010.

Einarsen, R.H. *A comparative study of volatility forecasting models* (2014):

Garman, M.B., and M.J. Klass "On the estimation of security price volatilities from historical data." *Journal of business* (1980):67-78.

Katsiampa, P. "Volatility estimation for Bitcoin: A comparison of GARCH models." *Economics Letters* 158 (2017):3-6.

Makridakis, S., E. Spiliotis, and V. Assimakopoulos "Statistical and Machine Learning forecasting methods: Concerns and ways forward." *PloS one* 13 (2018):e0194889.

McClelland, J.L., D.E. Rumelhart, and PDP Research Group "Parallel distributed processing." *Explorations in the Microstructure of Cognition* 2 (1986):216-71.

McNally, S., J. Roche, and S. Caton *Predicting the price of Bitcoin using Machine Learning*IEEE, 2018.

Robbins, H., and S. Monro "A stochastic approximation method." *The annals of mathematical statistics* (1951):400-7.

Schmidhuber, J. "Habilitation thesis: System modeling and optimization." *Page 150 ff demonstrates credit assignment across the equivalent of 1,200 layers in an unfolded RNN* (1993):

Wennström, A. "Volatility Forecasting Performance: Evaluation of GARCH type volatility models on Nordic equity indices." (2014):