Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/51/20241546

Cryptocurrency assets valuation based on statistics model and machine learning

Congyue Hao

Department of Statistics, King's College London, London NW3 3ES, the United Kingdom

k23037579@kcl.ac.uk

Abstract. As a matter of fact, cryptocurrency is an exchange media, which uses cryptographic functions to trade, which achieves decentralisation, transparency, and invariance using blockchain. With the rapid development since the first white book proposed in 2008. The market value of cryptocurrency has been exponentially expanded rapidly in recent years. On this basis, programmatic trading strategy are paid more attention by people in trading cryptocurrency contemporarily. With this in mind, this study provides three models, that is, Exponential Smoothing Model (ESM), Autoregressive Integrated Moving Average (ARIMA) model and Long short-term memory (LSTM) model to forecast the future price of cryptocurrency and use sum of square error to analyse the results of these prediction. According to the analysis, LSTM is the best model to forecast the price of cryptocurrency in this article. Overall, machine learning might be more useful to forecast in this fields than traditional statistical models. Investors and researchers in the future could continue to forecast the price of cryptocurrency combined with LSTM model.

Keywords: Cryptocurrency valuation, ESM, ARIMA, LSTM.

1. Introduction

Blockchain is a kind of digital ledger of economic and financial activities which could applied to document any transaction involving an item with inherent worth, not simply financial ones [1]. Each record is protected by the cryptographic principle and bounded in a same chain. The trade of cryptocurrency is processed according to a node-node network, which records the total historic information. Cryptocurrency is a part of blockchain, it is a kind of tokens used in this network. According to Doran [2], cryptocurrency is an exchange media, which uses cryptographic functions to trade. Cryptocurrency achieves decentralisation, transparency, and invariance using blockchain. One of the most features of cryptocurrency is that it is not affected by financial instituting and central institution [3, 4]. Nowadays, more than 4900 cryptocurrencies appears and 20000 cryptocurrency markets are established. One of the main cryptocurrency Bitcoin (BIT) was created in 2009, which is described as an electronic cash. Trading of cryptocurrencies is a behaviour to earn a profit from purchasing and selling cryptocurrencies. Generally, there exist two main kind of trading strategy: technical and fundamental strategy. Oberlechner suggests that most of traders are using these two approaches to trade. Recently, the third trading strategy, which one calls programmatic trading occurs [5].

Statistics model are applied in time series to study the trade of cryptocurrency. If there exists linear statistical relation between prices and variables, autoregressive moving average (ARMA) model could be used to forecast the price [6], Machine learning is an effective tool to forecast the price of cryptocurrency and trade [7]. In this field, LSTM model has been applied to forecast the price trend in the shares of the S&P 500. Fischer and Krauss argue that it might be probable to get significant information in financial markets with error noise [8]. Nevertheless, these models are not applied to forecast the price of cryptocurrencies directly.

Therefore, this article would use some statistical models and machine learning models to forecast the valuation of cryptocurrencies directly. Firstly, 5 years data of 10 different kind of cryptocurrencies will be chosen to forecast 1 year's price according to three models, which are Exponential Smoothing Model (ESM), Autoregressive Integrated Moving Average (ARIMA) Model and Long short-term memory (LSTM) Network. After that, sum of squared errors (SSE) will be calculated for the prediction of each model to identify the best model of them.

Table 1. Name and symbol of cryptocurrencies

Name	Symbol
Bitcoin	BTC
Litecoin	LTC
Ixcoin	IXC
Namecoin	NMC
Novacoin	NVC
Peercoin	PPC
Terracoin	TRC
Freicoin	FRC
Feathercoin	FTC
WorldCoin	WDC

2. Data and method

The price and volume of 10 different cryptocurrencies are grabbed according to the application programming interface (API) of CoinMarketCap from 30/08/2018 to 29/08/2023 by days. Types of cryptocurrencies are listed in Table 1. Close price and volume of the data are chosen to form the time series by days, which returns n = 1826 days' data. Three models will be used to forecast the value of cryptocurrencies for one year from 30/08/2023 to 29/08/2024. To test the accuracy of these models, 95% data will be taken as train sample to forecast 5% data from 31/05/2023 to 29/08/2023. Then, calculating the sum of square error (SSE) for different models to analyse these models.

Exponential Smoothing Model is a time series model suggested by Brown [9]. According to this model, the forecast is given by the weighted sum of observation at time t and the forecast value at time t. The formula is given by

$$F_{t+1} = \alpha x_t + (1 - \alpha)F_t \tag{1}$$

where F_{t+1} is the prediction, alpha is called the smoothing factor, $0 < \alpha < 1$, F1 = x0 with t = 0. In Exponential Smoothing Model, α is an unknown factor, so different value of α will be chosen to forecast to obtain the best prediction of this model. Since the observation values of the forecast year is unknown, the last observation of data will be fixed as xt to forecast Ft for the forecast year.

ARIMA model is a time series model introduced by Box et al. (2015) [10]. The autoregressive (AR) part is related to its own regression since the regressors can be considered as a shifted time series in the past, and the moving average (MA) part is the linear combination of errors for the past. The integrated (I) part indicates that original data value could be replaced with the higher order differences. Let {Xt} be a time series, then {Xt} is an ARMA(p,q) process if

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \cdots + \psi_q \varepsilon_{t-q} \tag{2}$$

where $\varepsilon_t \sim WN(0, \sigma^2)$, and the terms have no common factors. Let d be a integer, where d > 0, then {Xt} is an ARIMA(p,d,q) series if it satisfies:

$$\phi(B)(1-B)dX_t = \psi(B)\varepsilon_t \tag{3}$$

where B is the lag operator, ϕ_i is the parameter of AR part, ψ_i is the parameter of MA part. This model could eliminate the non-stationary of the time series by taking difference, and forecast the value of differential data.

n recent years, deep learning techniques have become increasingly important, and there are now more individuals or groups developing them. This fields aims to create algorithms to enable computers to learn without the need for explicit programming. When this study is modelling non-stationary variables and their complicated relationships in financial series, statistical and econometric approaches frequently create challenges [11]. Thankfully, deep learning methods are capable of recognising and handling these kinds of intricate systems [8]. Recurrent neural network (RNN) could used to collect historical information of the data to present the time series data. Nevertheless, traditional RNN model might encountered vanishing gradient problem due to repeated multiplication. Therefore, Hochreiter and Schmidhuber proposed a new RNN model called LSTM network, which could be applied to find the solution of vanishing gradient issue occurring in traditional RNN network [12]. Gate mechanism is introduced to LSTM model. Specifically, a forget gate (gt), a cell gate (ct), an input gate (it), and an output gate (ot) are included in an LSTM hidden layer. Fig 1 displays the structure of LSTM hidden unit layer [13].

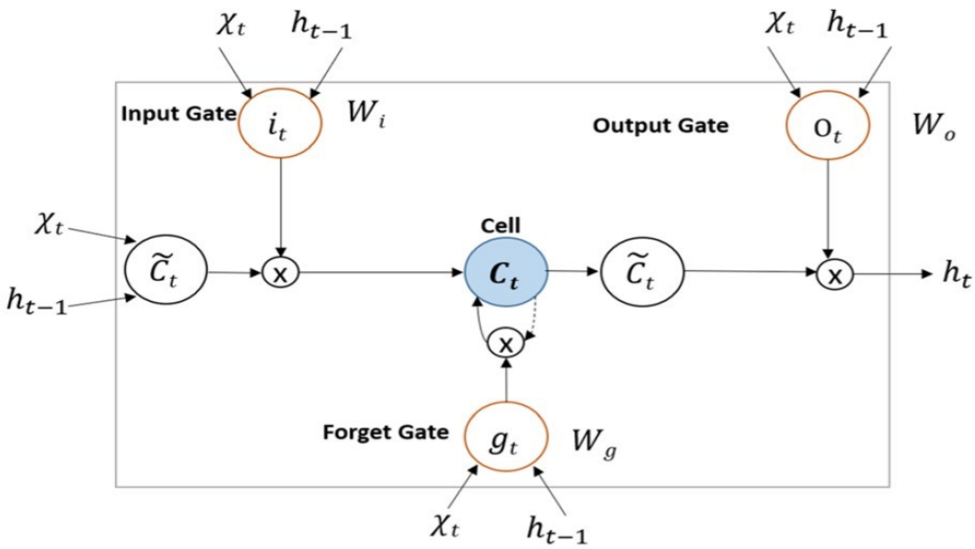


Figure 1. LSTM hidden unit layer [13].

Firstly, LSTM model should decide which information should be forgotten by forget gate g_t, which is dealt with the sigmoid function σ, i.e.,

$$g_t = \sigma(W_g \cdot [h_{t-1}, x_t] + b_g) \tag{4}$$

where Wg is the weight vector, b_g is the bias or threshold term. Then, new information will be added according to the input gate it, the formulas are given by

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{5}$$

and

$$C'_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_c) \tag{6}$$

DOI: 10.54254/2755-2721/51/20241546

the old cell information of the cell will be replaced as

$$C_t = g_t \times C_{t-1} + i_t \times C'_t \tag{7}$$

Finally, the output of this hidden layer will be

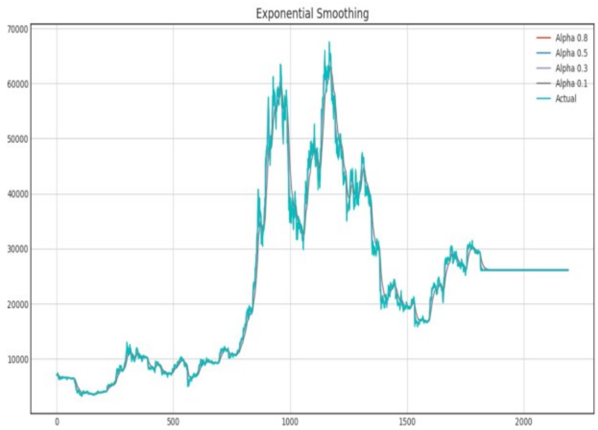
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{8}$$

and

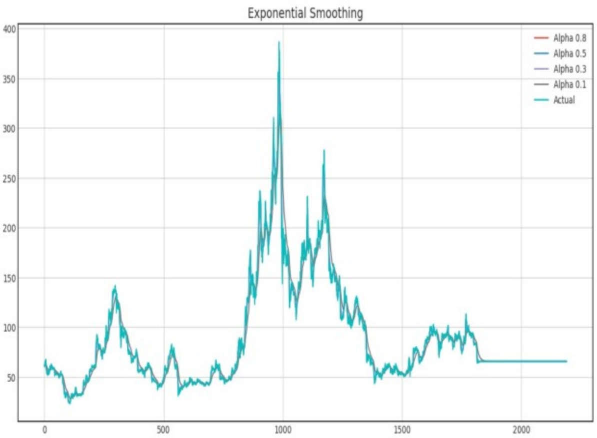
$$h_t = o_t \times \tanh(C_t) \tag{9}$$

3. Results and discussion

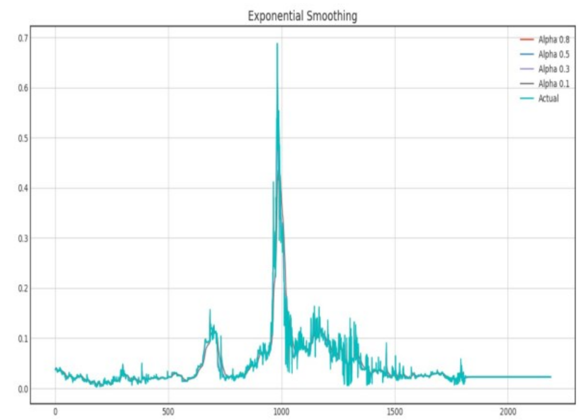
This section describes the forecast results of these three models and the square errors of tested data, i.e., last 5% data. Fig. 2 shows the forecast price plot of 10 cryptocurrencies described in Table 1 with different smoothing factor α = 0.1,0.3,0.5,0.8. According to these figures, the predicted value of price are convergent to a fixed value which is the value of last observation. More specifically, the sum of squared difference of tested data will be calculated for these 10 cryptocurrencies. Table 2 shows the sum of error squares these 10 cryptocurrencies. Generally, the lowest SSE occurs when α = 0.5,0.8.



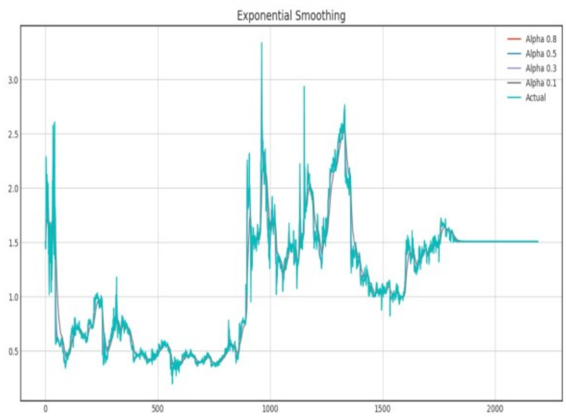
(a) Bitcoin's forecast price



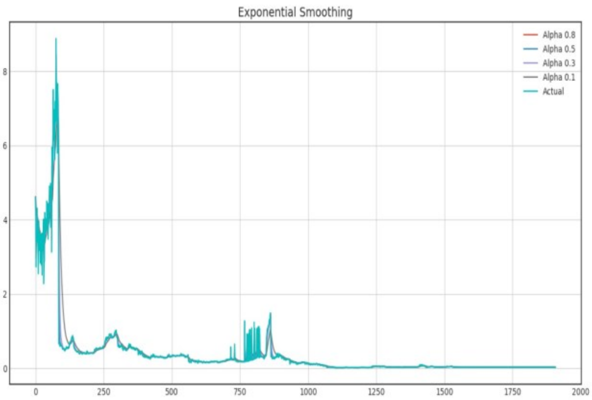
(b) Litecoin's forecast price



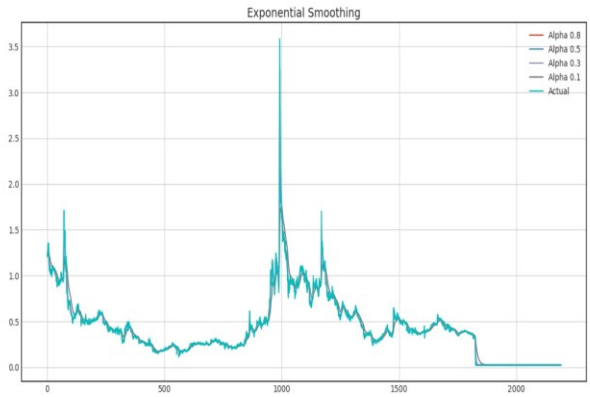
(c) Ixcoin's forecast price



(d) Namecoin's forecast price



(e) Novacoin's forecast price



(f) Peercoin's forecast price

Figure 2. Forecast price for 6 cryptocurrencies (Photo/Picture credit: Original).

According to the observation of these plots, the forecast value will converge to a fixed value by the formula due to the lack of the observed price value from 30/08/2023 to 29/08/2023. It means that exponential smoothing model might not fitted well to forecast the long-term price. Furthermore, the model's prediction has higher hysteresis quality when α close to 1. This character also could be found according to the formula. When the smoothing factor increases, the forecast value Ft+1 would be closer to the observation xt rather than the forecast value of Ft. By calculating the value of SSE, the SSE decrease with the increase of α when the cryptocurrencies are BIC, LTC, NMC, NVC, PPC, TRC, FRC and WDC. And the lowest SSE are taken when α = 0.5 when the cryptocurrencies are IXC and FTC. Therefore, the results of SSE could imply that when α is greater than 0.5, the value of forecast will be more accurate. Since the SSE is not high compared with the real values, this model might be suitable to forecast the short-term price value.

Table 2. Sum of square errors of tested data with α = 0.5

Symbols	α =0.1	α =0.3	α =0.5	α =0.8
BTC	212623807	85027823	47087424	28283890
LTC	8014	3561	2206	1499
IXC	0.007079	0.006197	0.005911	0.006382
NMC	0.47952	0.24939	0.20328	0.19958
NVC	0.004272	0.002091	0.001545	0.001431
PPC	0.05696	0.01544	0.008258	0.005768
TRC	2.1030e-04	7.6459e-05	5.4781e-05	5.3733e-05
FRC	6.2689e-05	4.2611e-05	3.3105e-05	2.6245e-05
FTC	1.1009e-04	8.1391e-05	7.6366e-05	8.2288e-05
WDC	0.004582	0.001078	0.0005845	0.0004321

Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/51/20241546

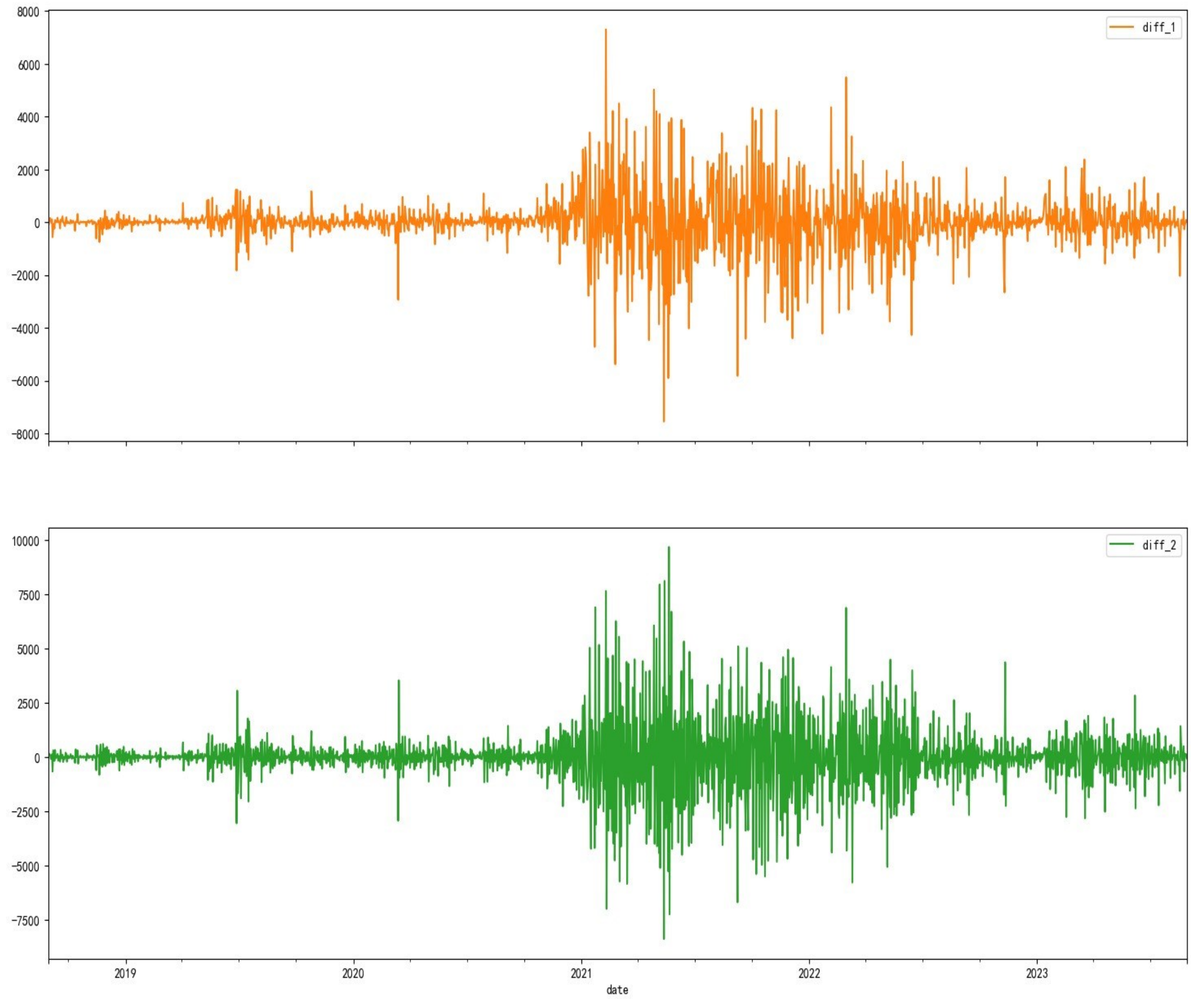
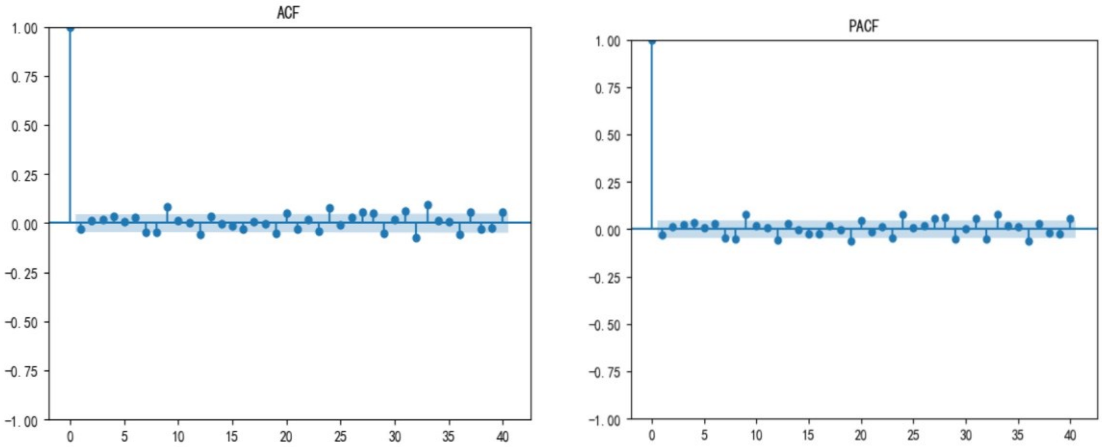


Figure 3. ARIMA difference of Bitcoin price (Photo/Picture credit: Original).

Figure 3 describes the 1st and 2nd difference of the Bitcoin price's data. According to this figure and adfuller test of the difference, the first order of difference is chosen to model the ARMA(p,q) model. Then ACF and PACF plot of the first difference are used to confirm the degree of p and q (seen from Fig. 4). By calculating the BIC in Fig. 5 of data and ACF and PACF plot, ARIMA(1,1,1) is confirmed to estimated the Bitcoin price. For different cryptocurrencies, ARIMA(1,1,1), ARIMA(2,1,3), ARIMA(8,1,8), ARIMA(8,1,8), ARIMA(1,1,1), ARIMA(1,1,1), ARIMA(8,1,8), ARIMA(2,1,1), ARIMA(1,1,1), ARIMA(1,1,1) are chosen separately to forecast the price value for next year. Table 3 describes the SSE of ARIMA model for these 10 cryptocurrencies. The results of BIC shows that p = 0 and q = 0 sometimes, which means that ARIMA model might not have clear autocorrelation and moving average structure. Some more complicated time series model such as seasonal autoregressive integrated moving average (SARIMA) model could be modelled to deal with potential seasonality factor. Since all of these cryptocurrencies are stable with the first order difference there exists linear trends in all of these price series, which means that some trend model might could be used to forecast the price of cryptocurrencies. The Fig. 6 shows that the forecast would have an approximately linear increase sometimes, which might not be an accurate prediction. By observing the SSE values, values of SSE are not excessively high, which means that the model still have certain accuracy.



(a) ACF (b) PACF

Figure 4. ACF and PACF plots of first differences for Bitcoin (Photo/Picture credit: Original).

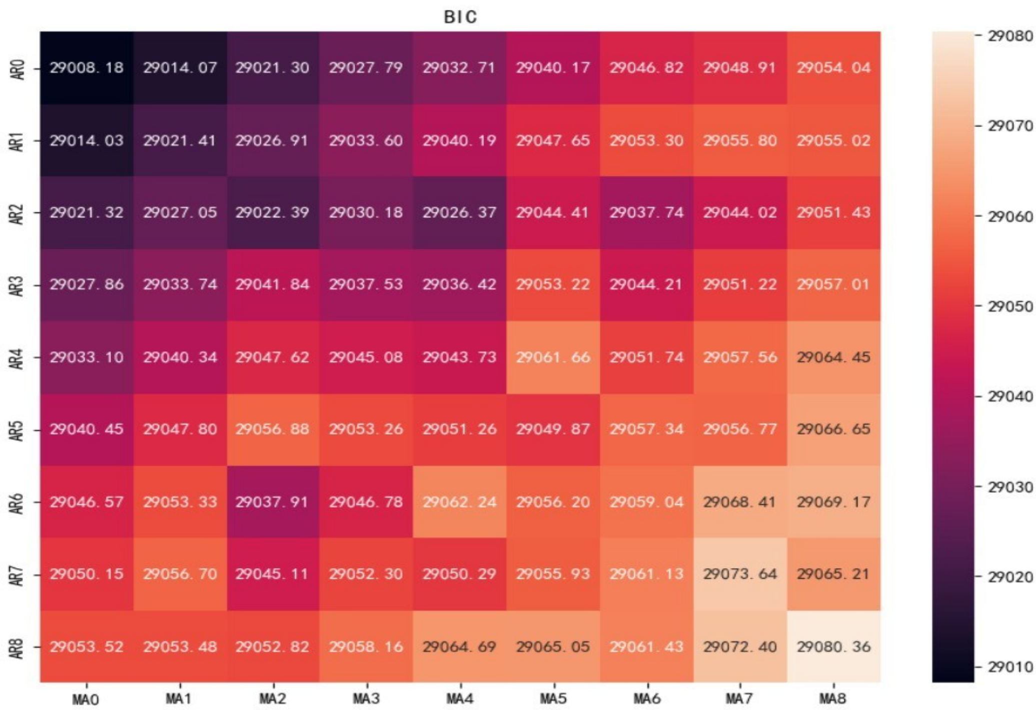


Figure 5. BIC of Bitcoin price (Photo/Picture credit: Original).

Table 3. SSE of ARIMA model

Symbols	ARIMA Forecast
BTC	25666102
LTC	1374
IXC	13.4384
NMC	0.1878
NVC	0.1655
PPC	12.1405
TRC	5.1e-05

Table 3: (continued).

FRC	4.197e-04
FTC	6.2266e-05
WDC	0.012527

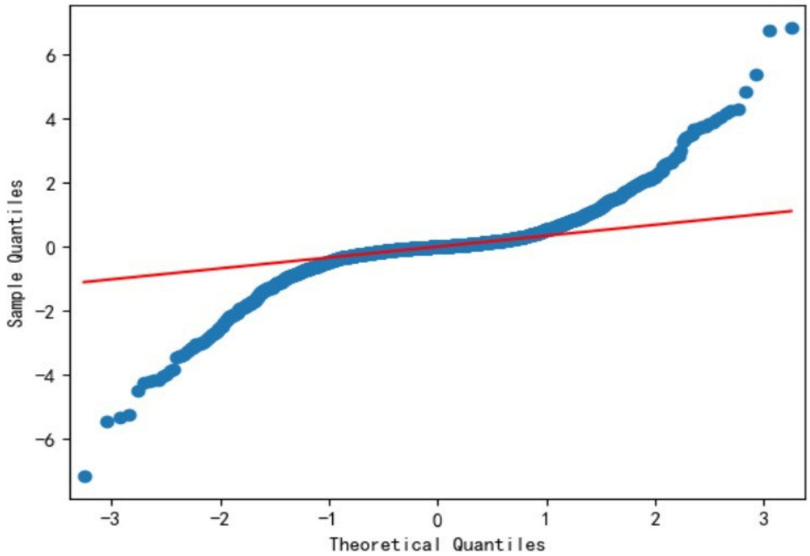
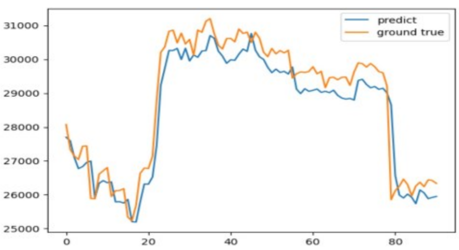


Figure 6. ARIMA residual plot of Bitcoin price (Photo/Picture credit: Original).

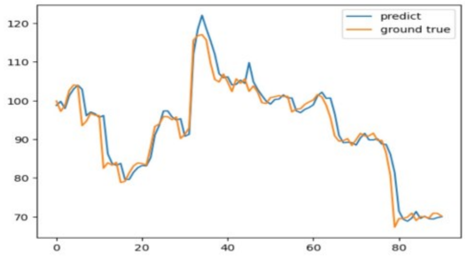
Table 4. SSE of ARIMA model

Symbols	LSTM Forecast
BTC	39760464
LTC	921.9137
IXC	0.0039447397
NMC	0.9725587
NVC	0.009949678
PPC	0.011220239
TRC	7.481491e-05
FRC	1.4276515e-05
FTC	6.413014e-06
WDC	0.0005164202

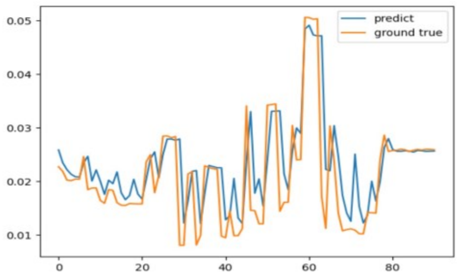
Figure 7 shows the test data prediction of cryptocurrencies from 31/05/2023 to 29/08/2023 with LSTM model and Table 4 describes the SSE of them. Figure 8 suggests the residuals plot of Bitcoin and Litecoin. The loss function are shown in Figure 9.



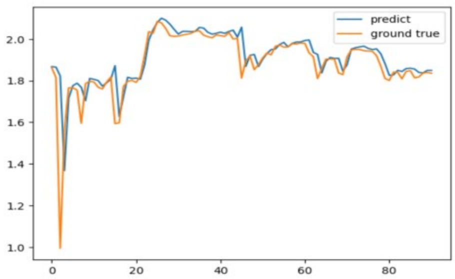
(a) BIT



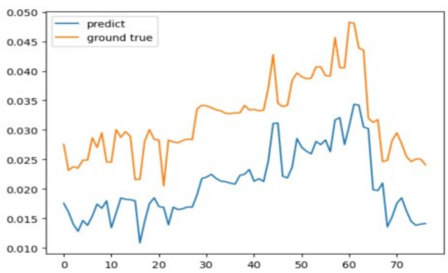
(b) LTC



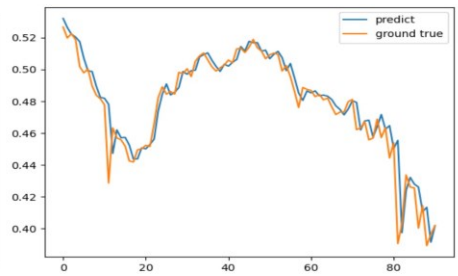
(c) IXC



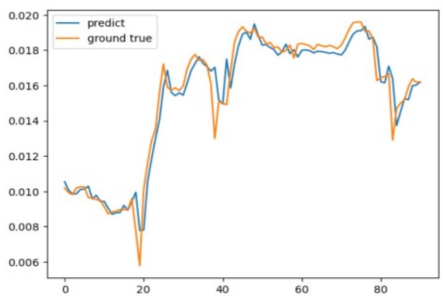
(d) NMC



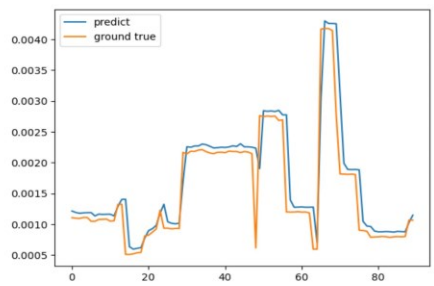
(e) NVC



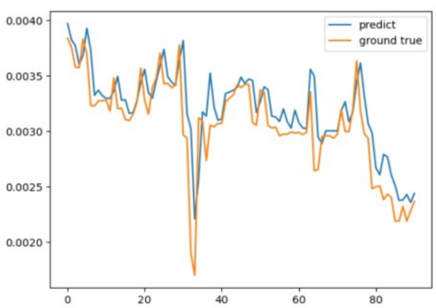
(f) PPC



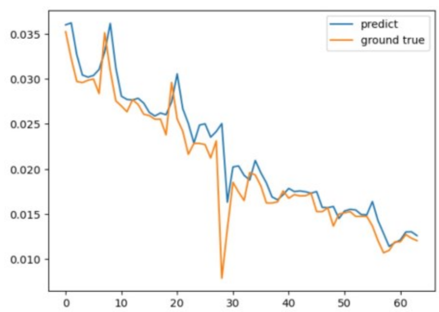
(g) TRC



(h) FRC

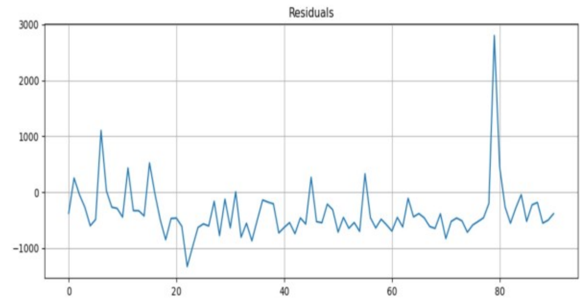


(i) FTC

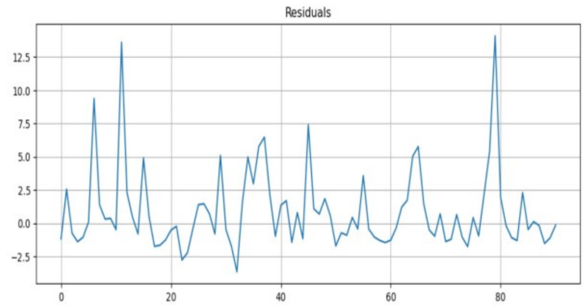


(j) WDC

Figure 7. LSTM forecast model for 10 cryptocurrencies (Photo/Picture credit: Original).



(a) BIT



(b) LTC

Figure 8. LSTM forecast model Residuals (Photo/Picture credit: Original).

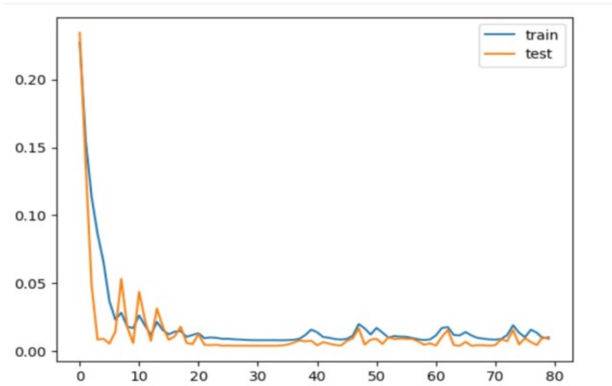


Figure 9. Loss function of Bitcoin price (Photo/Picture credit: Original).

The forecast plots shows that the fitting result are good for most of these cryptocurrencies prices except the prediction of NVC price. Figure 8 also suggests that residuals are near to the value 0, which could prove that this model is an accurate model for forecast. The values of SSE are in the lower range, which also shows LSTM model fits the data well. Furthermore, the variance of loss function of the test data revealed LSTM model does not overfitting. By comparing the SSE value of these three models, exponential smoothing model with $\alpha = 0.08$ has the lowest SSE for NVC, PPC and WDC; ARIMA model has the lowest SSE for BIC, NMC and TRC; LSTM model has the lowest SSE for LTC, IXC, FRC and FTC. Since the results of ARIMA model shows the price forecast might has a linear trend, which means exponential smoothing model could add trend factor to improve the accuracy of ESM. In addition, since the lowest SSE values of ESM model and ARIMA model are close to the value of LSTM model, LSTM model might be the best model of these three models to forecast the price of cryptocurrencies.

4. Limitations and prospects

Firstly, the data are obtained by days instead of by hours or minutes, which decrease the numbers of data and influence the prediction of models. This study does not analyse the effect of volume. However, trading volume is also an important factor of price forecast. Since volume reflects the activation of the cryptocurrency markets, and some models also need volumes as parameters to forecast. The exponential model might be simple to forecast these data, and the trend factor and seasonal factor are ignored in forecast. And this model is a linear model, which might not suitable to provide accurate results with unstable cryptocurrencies data. This model also not good to predict long-term results due to the lack of ability to capture long-term trend. ARIMA model is sensitive to the data mutation and outlier. Since cryptocurrencies price has data mutation frequently, the accuracy of ARIMA model might be affected. Only 80 neuron layers are set to train the LSTM model, which might not be enough to contribute a good

DOI: 10.54254/2755-2721/51/20241546

LSTM model. This article does not analyse the output of hidden layer unit, which might cause the significance of different features are unknown. This article only focuses on the cryptocurrencies assets valuation, but the risk analysis and asset allocation of cryptocurrencies market also are important in this field. Study in the future could explore more models and methods to analyse these projects according to statistics and finance tools. Future work could focus on some more complicated model, such as ARIMA-LSTM model, SARIMA model, GARCH model and so on. Volatility could be analysed to describe the trend of the price of cryptocurrencies. More features such as trading volume, similar cryptocurrencies price, price of consecutive days or hours could be added to contribute the model. In addition, the market sentiment of cryptocurrency has essential influence to the price. Emotion Index model could be considered to establish to forecast the effects of market sentiment. Macroeconomics elements also affect the price of cryptocurrencies. Inflation rates, interest rates and policies should be added to build macroeconomics models, which might be useful to predict the price in the future more accurate. Finally, it is interesting to find more effective models of machine learning to model the price of cryptocurrencies.

5. Conclusion

In conclusion, this study provides three models to forecast the price of 10 cryptocurrencies, which are exponential smoothing model, ARIMA model and LSTM model. According to the modelling, the prediction plots and residual plots are provided to show the results directly. And the sum of square errors is calculated for each model to compare with each other. LSTM model provides the best prediction results of them. Due to the lack of the trend element, exponential smoothing model could not provide well results in the end. Since this article focus on the prediction of values, risk analysis and asset allocation of cryptocurrencies market does not mention, and many features such as trading volume, inflation rates are not considered. Work in the future could develop more helpful models to solve this problem and forecast the price of cryptocurrencies. This article shows that LSTM model is effective to forecast the price of cryptocurrencies. People in the future could use this model combined with other model to explore more accurate approaches to predict the price of cryptocurrencies.

References

[1] Tapscott D and Tapscott A 2016 Blockchain revolution: how the technology behind bitcoin is changing money business and the world Penguin vol 11.

[2] Doran M D 2014 A forensic look at bitcoin cryptocurrency Doctoral dissertation of Utica College.

[3] Harwick C 2016 Cryptocurrency and the problem of intermediation The Independent Review vol 20(4) pp 569-588.

[4] Rose C 2015 The evolution of digital currencies: Bitcoin a cryptocurrency causing a monetary revolution International Business and Economics Research Journal (IBER) vol 14(4) pp 617-622.

[5] Oberlechner T 2001 Importance of technical and fundamental analysis in the European foreign exchange market International Journal of Finance Economics vol 6(1) pp 81-93.

[6] Choi B 2012 ARMA model identification Springer Science Business Media vol 7.

[7] McNally S, Roche J and Caton S 2018 Predicting the price of bitcoin using machine learning In 2018 26th euromicro international conference on parallel distributed and network-based processing (PDP) pp 339-343.

[8] Fischer T and Krauss C 2018 Deep learning with long short-term memory networks for financial market predictions European Journal of Operational Research vol 270(2) pp 654–669.

[9] Brown R G 1956 Exponential smoothing for predicting demand Little vol 16.

[10] Box G E, Jenkins G M, Reinsel G C and Ljung G M 2015 Time series analysis: forecasting and control John Wiley and Sons vol 3.

[11] Dixon M, Klabjan D and Bang J H 2015 Implementing deep neural networks for financial market prediction on the intel xeon phi Proceedings of the 8th workshop on high performance computational finance pp 1–6.

DOI: 10.54254/2755-2721/51/20241546

[12] Hochreiter S and Schmidhuber J 1997 Long short-term memory Neural computation vol 9(8) pp 1735-1780.
[13] Garc´ıa-Medina A and Luu Duc Huynh T 2021 What drives bitcoin? An approach from continuous local transfer entropy and deep learning classification models Entropy vol 23(12) p 1582.