



Linkage Analysis Between Bitcoin and Nasdaq Index Based on ARIMAX Model

Ruixin Hu and Xuecheng Wang^(✉)

School of Economics and Management, North China University of Technology, Beijing, China
wxc@outlook.com

Abstract. Bitcoin is currently the most widely used encryption currency in the world, and the Nasdaq Index, as the world's first stock market to use electronic trading, has a certain impact on the price of Bitcoin. Based on the Bitcoin closing price and Nasdaq index data from January 2020 to May 2022, this paper predicts the price of Bitcoin by using ARIMA and ARIMAX models respectively. The linkage was confirmed by the correlation test, and the fitting and prediction effect of the ARIMAX model with the Nasdaq index as the input variable were better than the ARIMA model.

Keywords: ARIMAX model · Bitcoin · Financial data · forecast

1 Introduction

With the continuous progress of technology and the continuous development of digital economy, money is gradually becoming virtualization and digitalization. As of December 31, 2021, there were 8,153 cryptocurrencies worldwide, with a total market value of \$2.1 trillion according to Investing data. Among them, Bitcoin ranked first both in terms of market share and market trading volume, with a market value of \$800 billion, accounting for more than 40%. The blockchain-based native token Bitcoin is a private digital currency issued by the private sector which based on commercial credit. Over the past decade, the price of Bitcoin has risen from a few cents to more than \$50,000. Cryptocurrency advocates regard Bitcoin as digital gold: Firstly, Bitcoin is currently the most valuable digital currency in the market, and the originator of digital currency; Secondly, Bitcoin is limited, 21 million total, extremely scarce and not issued by a single institution or country; Finally, Bitcoin is known and recognized as having value, and nowadays can be used in many stores purchasing and transactions in the market.

The Nasdaq Index, also known as the U. S. Technology Index, has listed companies covering all the new technology industries, mainly composed of hundreds of the fastest growing advanced technology, telecommunications and biological companies in the United States, mainly representing the emerging technology industries. It is also the first stock market in the world to use electronic trading. Therefore, by exploring the connection between Bitcoin and Nasdaq index, we can observe the impact of Nasdaq index on the price volatility of Bitcoin, and provide some theoretical reference for the prediction of Bitcoin price, which has certain practical significance.

© The Author(s) 2022

V. Balakrishnan et al. (Eds.): ICMEIM 2022, AHSSEH 6, pp. 905–912, 2022.

https://doi.org/10.2991/978-94-6463-044-2_113

2 Literature Review

Yermack [1] (2014) believes that Bitcoin lacks many attributes of a useful currency and is more represented as an investment commodity, while there is no correlation between the dollar price and the currency rate of the dollar against other major currencies, nor with the gold price. Selgin [2] (2015), collectively, these cryptocurrencies generally referred to as the "synthetic commodity currencies," which is contrary to legal tender: they only change their production share, not total coin production. At the same time, it is found that some properties of Bitcoin and gold are relatively similar to the scarcity performance, and there is no intrinsic value like the general currency, like the US dollar. Lian Cheng [3] (2019), the emergence of digital currency does not pose a "new" substantive challenge to the existing mainstream monetary theory, but it does reveal the need to reflect on the traditional conceptual framework and expand the analysis perspective.

Because the price of Bitcoin has a large volatility and nonlinear change characteristics, so the prediction of Bitcoin price has been a problem worth studying. Gege Duan [4] (2021) establishes the ARIMA model directly based on the automatically generated ARIMA model to select the best model for prediction. Ning Zhang et al [5] (2021) built four hybrid prediction models based on the long and short-time memory network, using LSTM to predict the reconstructed subsequences respectively, and finally superimposed them to obtain the final prediction results, and found that the prediction accuracy of the hybrid model was better than that of a single model.

In the above model, the impact of other variables on Bitcoin price is not considered in the Bitcoin price prediction, but in fact, many studies have pointed out that a variety of factors have a certain impact on the closing price of Bitcoin. Kristoufek [6] (2015) studies the potential drivers of the price of Bitcoin and further studies the potential impact of the Chinese market and believes that trade transactions, money supply and price levels play a long-term role in the price of Bitcoin. Xiaofeng Liu [7] (2020) established the VAR model, analyzed the price influencing factors of digital currency represented by Bitcoin, and made the pulse analysis and variance test of these influencing factors, and summarized the factors affecting the price of Bitcoin. Finally, the significance of the empirical research and theoretical research of digital virtual currency is put forward.

From the above research literature, we can conclude that although the nature of Bitcoin is still controversial, it is certain that Bitcoin owns both standard financial assets and speculative assets. At the same time, after years of development, the Bitcoin market is also improving. Its price is not only driven by policy, but also influenced by macro financial indicators.

3 Theoretical Models and Methods

3.1 ARIMAX Model

The ARIMAX model is an extended version of the ARIMA model. Because the establishment of the ARIMA model is not complicated, only the analysis and prediction of the univariates do not require the help of other exogenous variables, so the results obtained by establishing this model are usually not accurate. If other variables are included as influencing factors, the model will be better.

The ARIMAX model, also known as the dynamic regression model, is one of the time-series prediction and analysis methods. In ARIMAX (p, d, q), AR is "autoregressive", p is the number of autoregressive items; MA is "sliding average", q is the number of sliding average items, d is the difference number (order) made to become a stationary sequence; X is the introduced exogenous variable. The basic idea of this model is to assume that the response sequence $\{Y_t\}$ And the input sequence $\{X_{1t}\}, \{X_{2t}\}, \dots, \{X_{nt}\}$ Both are smooth sequence, and you can build a regression model between the two:

$$Y_t = \mu + \sum_{i=1}^k \frac{\theta_i(B)}{\Theta_i(B)} B^{li} x_k + \varepsilon_t \tag{1}$$

Here: μ is the constant; B is the delay operator; $\theta_i(B)$ and $\Theta_i(B)$ are the polynomial of moving average coefficient of the *i*th input variable and the polynomial of autoregressive coefficient of the *i*th input variable, respectively; $\{\varepsilon_t\}$ is the residual sequence.

At the same time, because the output sequence $\{Y_t\}$ and the input sequence $\{X_{1t}\}, \{X_{2t}\}, \dots, \{X_{nt}\}$ are both stable sequences, that is, the linear combination of the stationary sequences is still stable, so $\{\varepsilon_t\}$ is also stable:

$$\varepsilon_t = Y_t - (\mu + \sum_{i=1}^k \frac{\theta_i(B)}{\Theta_i(B)} B^{li} x_k) \tag{2}$$

The residual sequence $\{\varepsilon_t\}$ information was extracted by the ARMAX model, and finally jointly get:

$$\begin{cases} Y_t = \mu + \sum_{i=1}^k \frac{\theta_i(B)}{\Theta_i(B)} B^{li} x_k + \varepsilon_t \\ \varepsilon_t = \frac{\psi_i(B)}{\gamma_i(B)} a_t \end{cases} \tag{3}$$

Here: $\psi_i(B)$ and $\gamma_i(B)$ $a_i(B)$ the autoregressive coefficient polynomial of the residual series and the moving average coefficient polynomial of the residual series; a_t is a pure random sequence. Since the premise mentioned above for establishing the ARIMAX model is that both the output sequence and the input sequence are stable, so once the condition model is not satisfied, it is untenable, that is the false regression. Therefore, we should conduct the stationarity test to avoid this situation.

4 Empirical Analysis

4.1 Data Sources

As the Chinese brand of Investing.com, the world’s fourth largest financial website, It provides the real-time market quotation and news and information of hundreds of thousands of financial investment products, including global stocks, foreign exchange, futures, bonds, funds and digital currencies, as well as a variety of investment tools. This article uses the Bitcoin price and Nasdaq index provided by Investing.com, and selects the data from January 2020 to May 2022 for study, and draws the timing chart of Nasdaq index and Bitcoin closing price, which is shown in Fig. 1. It is also logarithmic normalized.

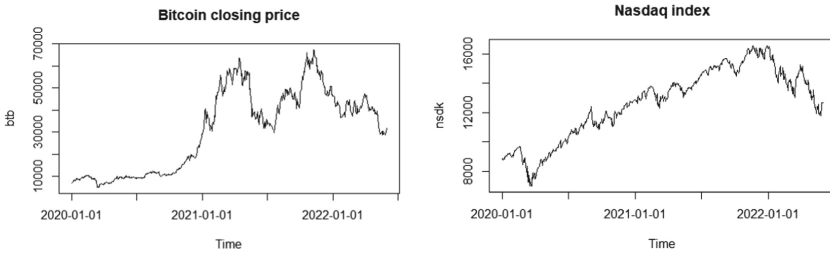


Fig. 1. The Bitcoin closing price and the Nasdaq index timing chart [Owner draw]

Table 1. Descriptive statistics

Variables	Bitcoin Price	Nasdaq index	Bitcoin (Log yield)	Nasdaq index (Log yield)
Sample size	608	608	607	607
Minimum value	4826	6994.29	-0.497	-0.130
Maximum value	67527.9	16573.34	0.192	0.096
Average	31004.6875	12655.8240	0.0023	0.0006
Variance	3.407169e + 08	5.754827e + 06	0.00250	0.00036
1. Quartile	10595.02	11034.77	-0.018	-0.0006
Standard deviation	18443.33	2396.945	0.049	0.019
Skewness	0.048	-0.415	-1.793	-0.665
Kurtosis	-1.43	-0.83	18.634	6.448

4.2 Descriptive Statistics

Table 1 shows the descriptive statistics for the Bitcoin and Nasdaq indexes. From this, we can find that the difference between the Bitcoin closing price and the Nasdaq index is both large, and the standard deviation is also large, indicating that the two variables fluctuated wildly during this period. Both Bitcoin and Nasdaq are smaller in the upper quartile and the variance is close to zero, less discrete, so they are not dispersed, largely weakening the variance compared to the original data.

4.3 Stability Test

According to the ADF unit root test, the T statistic is at 5%, and the Bitcoin and Nasdaq index p-values are 0.902 and 0.9847, respectively, both greater than 0.05. The null hypothesis H0 is not rejected to further verify that both sequences are non-stationary time series.

Table 2. The ADF test

ADF Inspection	p-value	Conclusion
Bitcoin price	0.902	Don't refuse
The logarithmic yield on Bitcoin's closing price	<0.01	refuse
Nasdaq index	0.9847	Don't refuse
The log yield of the Nasdaq index	<0.01	refuse

We get the logarithmic sequence through the `ndiff()` function in R language to make the first-order difference. As shown in Table 2, it is less than 0.05, indicating that the logarithmic first-order difference sequence is stationary sequences.

4.4 Estimation of the ARIMAX Model

After the stationarity test of input variables and response variables was carried out and passed, ARIMAX model with input variable regression was constructed to fit the changes of response series. Through the R language of the `arima()` function software algorithm automatically determines p , d , and q as 3,0,3, respectively, or ARIMAX (3,0,3). After processing the data, the two models were fitted separately, and the "maximum likelihood (ML)" and "minimum sum of square (CSS)" and the above two hybrid methods "CSS-ML" were used to obtain the regression results.

The best-fit model was also selected according to the data obtained by the Akaike information criterion (AIC). According to the results, the AIC of the ARIMAX model is significantly smaller than the AIC value of the ARIMA model, indicating that the model effect of the ARIMAX is better than the ARIMA model.

The regression output from Table 3 can be obtained based on the hybrid method "CSS-ML" model of ML and CSS over the other two methods in both p-value significance and AIC value, so we chose this method to predict the two models.

4.5 Model Prediction Results

Once the most suitable model has been selected for the time-series data, the parameters of the model can be used as a predictive model to predict the future values of the time-series. The prediction was made using the `forecast()` function according to the results of the model fitting function.

The ARIMA and ARIMAX models were used to predict the known Bitcoin closing price between January 2020 and May 2022 and compare them with the actual values. The prediction results and prediction errors are shown in Table 4.

By comparing the predicted results and their errors, it was found that the ARIMAX model combining the Nasdaq index as an exogenous variable predicted better than the ARIMA model, with an average error of around 2.47%. Therefore, it can be concluded that the prediction accuracy of the ARIMAX model by introducing exogenous variables is relatively high, which is more conducive to the prediction of Bitcoin price.

Table 3. Regression Results output

Variable	ARIMA			ARIMAX		
	Model1 (3,0,3) “ML”	Model2 (3,0,3) “CSS”	Model3 (3,0,3) “CSS-ML”	Model4 (3,0,3) “ML”	Model5 (3,0,3) “CSS”	Model6 (3,0,3) “CSS-ML”
ar1	-0.601 (0.188)	0.561** (0.021)	0.814*** (0.00001)	0.215 (0.755)	0.468** (0.038)	0.259 (0.691)
ar2	-0.889*** (0.000)	-0.785*** (0.000)	-0.755*** (0.000)	-0.800*** (0.000)	-0.802*** (0.000)	-0.799*** (0.000)
ar3	-0.441 (0.238)	0.558*** (0.007)	0.769*** (0.0000)	0.257 (0.662)	0.450** (0.024)	0.294 (0.593)
ma1	0.496 (0.274)	-0.636*** (0.010)	-0.891*** (0.00001)	-0.268 (0.700)	-0.530** (0.026)	-0.313 (0.636)
ma2	0.901*** (0.000)	0.901*** (0.000)	0.894*** (0.000)	0.844*** (0.000)	0.871*** (0.000)	0.847*** (0.000)
ma3	0.451 (0.256)	-0.576*** (0.009)	-0.801*** (0.0000)	-0.223 (0.714)	-0.437** (0.045)	-0.261 (0.648)
intercept	0.002 (0.185)	0.002 (0.282)	0.003 (0.244)	0.002 (0.292)	0.002 (0.367)	0.002 (0.295)
xreg				0.993***	0.995***	0.993***
sigma2	0.002	0.002	0.002	0.002	0.002	0.002
Note	*p < 0.1	**p < 0.05	***p < 0.01	*p < 0.1	**p < 0.05	***p < 0.01
AIC	-1954.43		-1955.07	-2049.08		-2049.08

Table 4. The ARIMA and ARIMAX prediction results

time	ARIMA			ARIMAX		
	predicted value	Corresponding closing price	average error	predicted value	Corresponding closing price	average error
2022.6.1	-0.0065	31323.51	7.32%	-0.0146	30744.30	2.47%
2022.6.2	0.0040	31612.26		-0.0029	30538.47	
2022.6.3	0.0098	31938.11		0.0264	32451.33	
2022.6.6	0.0046	32669.39		-0.0120	31569.77	
2022.6.7	-0.0035	32408.02		-0.0097	30869.96	
2022.6.8	0.0048	32770.33		0.0062	31314.18	
2022.6.9	0.0073	33327.70		-0.0142	30306.06	
2022.6.10	0.0083	33334.07		-0.0233	28724.44	

5 Conclusion

This paper started with a descriptive statistical analysis of the Nasdaq index and Bitcoin closing price and its logarithmic yield, and found that the two variables fluctuated violently during this period. At the same time, because the timing graphs of the variables are all non-stationary sequence, the first-order difference of the logarithmic data is required, and the resulting sequence of log yields is stable. On this basis, we have established ARIMA (3,0,3) model and ARIMAX model with Nasdaq index as the input variable respectively, and estimated the two models for the optimized models. By comparing the prediction results and the average error of the two models, we have found that the prediction accuracy of ARIMAX model is higher than that of ARIMA model, which is more suitable for the prediction of the closing price of Bitcoin.

Thus, compared with the single time series model, the introduction of the Nasdaq index multiplex time series model can improve the reliability of the prediction results, which once again shows that the Nasdaq index has a great impact on the price of Bitcoin, and there is a linkage between the two.

As a digital currency, Bitcoin is highly volatile and non-linear, and its price is affected by many factors, such as investors' speculative behavior, and various macro policy factors, which may have a large impact on the price of Bitcoin. In this paper, we only introduce the Nasdaq index as an influencing factor to study the time series of Bitcoin's closing price. There are few variables, and the accuracy of the prediction results is still not enough. So after the prediction of Bitcoin price can be introduced into multiple time variables, including stock trading index, exchange rate, trade trading, etc., multi-angle further study of Bitcoin price, at the same time, also can be the model optimization, such as Arima-Transformer model in the combination model, Coupla-ARIMA model to predict the Bitcoin price, to further improve the prediction effect of the model.

In short, whether as an investor or as a researcher, multi-dimensional and multi-faceted research methods must be conducive to improving the accuracy of Bitcoin price prediction, so we should make better use of these methods for practical analysis to further improve the theoretical model.

References

1. Yermack, David. Bitcoin Economics [J]. *Technology Review*, 2014, 117(2).
2. George Selgin. Synthetic commodity money[J]. *Journal of Financial Stability*, 2015, 17.
3. Cheng Lian. Digital currency: from economy to society [J]. *Social Science Front*, 2020 (06): 60-72.
4. Duan Gege. Research on Bitcoin Price Prediction Based on the ARIMA Model [J]. *Modern Marketing (Next ten days)*, 2021 (01): 27-29.
5. Zhang Ning, Fang Jingwen, Zhao Yuxuan. Bitcoin price prediction based on the LSTM hybrid model [J]. *Computer Science*, 2021, 48 (S2): 39-45.
6. Kristoufek Ladislav. What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis.[J]. *PloS one*, 2015, 10(4).
7. Liu Xiaofeng. An Empirical Study on the Impact Factors of Digital Currency Price [D]. Shandong University, 2020.
8. Yao Xiuxia, Du Jie. Application of ARIMAX model in birth rate prediction of population in Province A [J]. *Hengyang Normal College Journal*, 2021, 42 (03): 33-39

9. Zhu Junrong, Huang Ailian. Linkage Analysis of RMB Exchange rate and China's inbound tourism income The —— is based on the ARIMAX model [J]. Journal of Qiqihar University (Natural Science Edition), 2021,37 (02): 88-94
10. Hossain Md Sabbir, Ahmed Sulaiman, Uddin Md Jamal. Impact of weather on COVID-19 transmission in south Asian countries: An application of the ARIMAX model.[J].The Science of the total environment,2020,761.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

