Research on the Correlation between Digital Cryptocurrencies Based on Spss Statistical Method

Hongze Qiu

E-mail: 3381832227@qq.com

School of Business, East China University of Science and Technology, Shanghai 200237, China

Abstract. In this paper, SPSS software was used to study the correlation of digital cryptocurrencies in the context of the epidemic, and the conclusion that "there is significant correlation between various digital cryptocurrencies" was reached. Thus, it provides reference value for the analysis of the correlation between the price collapse of digital cryptocurrency under the epidemic and the price recovery of digital cryptocurrency after the epidemic.

Keywords: Digital Cryptocurrency, Correlation, SPSS, Cluster Analysis, Regression Analysis

1 Introduction

Since 2020, when the COVID-19 has had a huge impact on the global economy, although Bitcoin has risen sharply in the violent shock, and hit the record high price of $64863 in April 2021, it then plummeted all the way to $25401.29 on May 12[1]. At the worst, it plummeted by $4.5 billion an hour, leaving 470000 people without money. Coincidentally, other digital cryptocurrencies such as Ethereum also fell sharply at the same time[2].

At present, scholars' research on the correlation of digital cryptocurrency is not complete under the circumstances of dramatic changes in the epidemic environment[3]. In order to make contributions to the prediction of the correlation of the price recovery of digital currency after the epidemic, this paper will use the data of digital cryptocurrency during the period of relatively large rise and fall in the past three years to analyze based on SPSS software. At the same time, it is assumed that there is a significant correlation between various digital cryptocurrencies[4].

2 Methods

2.1 Data Collection and Sorting

This paper selects the historical daily rise and fall data of Bitcoin, Ethereum, Teda, USD Coin, BNB, XRP, Binance USD, Ida, Solana, and Dog Coin, which are the top ten digital cryptocurrencies in the comprehensive ranking from March 1, 2022 to April 30, 2022, as samples for research. This paper takes the daily fluctuation data of 10 kinds of digital cryptocurrencies from March 1, 2022 to April 30, 2022 as the research sample, and takes time and the type of digital cryptocurrencies as the variables, a total of 11 variables.

ICBBEM 2023, May 19-21, Hangzhou, People's Republic of China

DOI 10.4108/eai.19-5-2023.2334355

The sample variance model, covariance model and correlation coefficient model are as follows[5]:

$$r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj}}\sqrt{s_{kk}}} = \frac{\sum_{i=1}^{n}(x_{ij}-\overline{x_j})(x_{ik}-\bar{x}_k)}{\sqrt{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}\sqrt{\sum_{k=1}^{n}(x_{ik}-\overline{x_k})^2}} \tag{1}$$

i=1, 2, …, p; j=1, 2, …, p; k=1, 2, …, p. In addition, for all j and k, $r_{jk} = r_{kj}$.

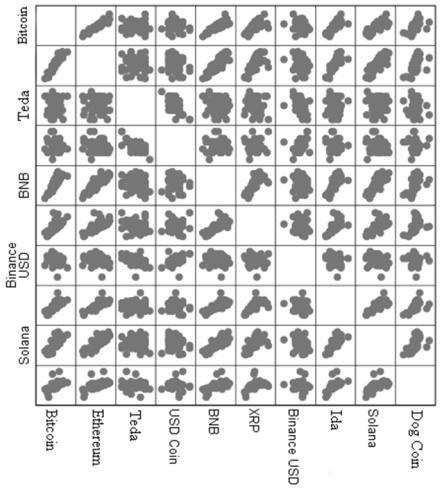

Fig. 1. Scatter Chart Based on Statistical Data

As shown in Figure 1, there is a positive correlation between most digital cryptocurrencies, so it is preliminarily predicted that there is a certain correlation between various digital cryptocurrencies[6].

2.2 Cluster Analysis

In this paper, the method of systematic cluster analysis is used. In cluster analysis, the concept of distance is selected to describe the "proximity" between samples; The similarity of indicators is characterized by similarity coefficient; Qualitative variables are converted into numerical variables to facilitate research[7].

For the similarity between the two indicators, we use the similarity coefficient C_{ij} is used to measure the distance between samples, and the average square distance between two types of elements is used to define the square of the distance between the two types[8]:

$$d_{ij} = [\sum_{k=1}^{p}(x_{ik}-x_{jk})^2]^{\frac{1}{2}},\ D_{kr}^2 = \frac{n_p}{n_r}D_{kp}^2 + \frac{n_q}{n_r}D_{kp}^2 \tag{2}$$

Use SPSS software to get the tree diagram, as shown in Figure 2.

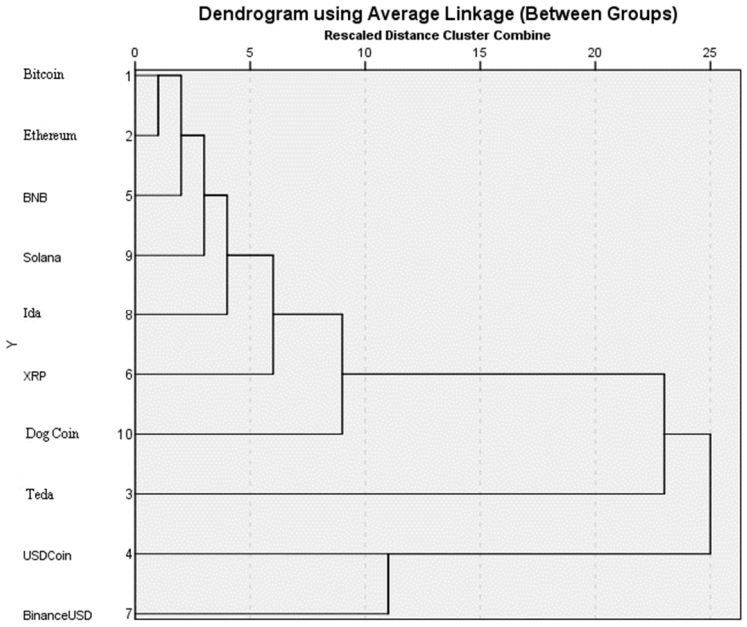

Fig. 2. Dendrogram Formed by Cluster Analysis

2.3 Univariate Linear Regression -- Taking Bitcoin and Ethereum as Examples

Variable selection: study the impact of daily rise and fall of Bitcoin on the daily rise and fall of Ethereum, select the daily rise and fall of Bitcoin as the explanatory variable, and the daily rise and fall of Ethereum as the explanatory variable.

Model establishment: x represents the daily rise and fall of Bitcoin, and y represents the daily rise and fall of Ethereum[9]. Let x and y satisfy a linear equation of one variable, with $\hat{\beta}_0,\hat{\beta}_1$ represents respectively β_0,β_1, then the univariate linear empirical regression equation of y with respect to x is:

$$y = \beta_0 + \beta_1 x + \varepsilon,\ \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \tag{3}$$

In order to obtain regression parameters from sample data β_0 and β_1, we need to use ordinary least squares estimation, regression coefficient β_0,β_1Estimated value of $\hat{\beta}_0,\hat{\beta}_1$ Meet:

$$Q(\hat{\beta}_0,\hat{\beta}_1) = \min_{\beta_0,\beta_1} \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2, \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \tag{4}$$

Solve the normal equation to obtain the least squares estimates of β_0,β_1:

$$\hat{\beta} = \bar{y} - \hat{\beta}_1 \bar{x},\ \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{5}$$

The least squares estimation should meet the following requirements at the same time: the estimators $\hat{\beta}_0$,$\hat{\beta}_1$ are linear functions of random variables y_i; E($\hat{\beta}_1$)=β_1, E($\hat{\beta}_0$)=β_0;The variance of $\hat{\beta}_0$ is $(\frac{1}{n}+\frac{\bar{x}^2}{L_{xx}})\sigma^2$.

F test is used to judge whether the empirical regression equation really describes the statistical law between variables y and x. The original hypothesis of the test is $H_0: \beta_1 = 0$, and its opposite hypothesis is $H_1: \beta_1 \neq 0$. F test is based on the decomposition formula of the sum of squares:

$$S_T = S_R + S_E \tag{6}$$

The significance of the regression equation is tested directly from the regression effect. The greater the S_R, the better the regression effect[10].

Under the normal assumption, F follows the F distribution with (1, n-2)d degree of freedom. When the F value is greater than the critical value Fa(1, n-2), H0 is rejected, indicating that the regression equation is significant, and there is a significant linear relationship between x and y.

The results of univariate linear regression are shown in Figure 3.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	.119	.149		.802	.426	-.178	.417
	Bitcoin	1.001	.051	.931	19.638	.000	.899	1.103

a. Dependent Variable: Ethereum

Fig. 3. Results of Univariate Linear Regression

3 Results

3.1 Interpretation of Statistical Results

Mean value: the sample mean value of seven digital cryptocurrencies is negative, the sample mean value of one digital cryptocurrency is zero, and the sample mean value of two digital cryptocurrencies is close to zero. From this, through simple mean value analysis, it can be seen that the overall daily rise and fall of digital cryptocurrency in the same period has the same downward trend.

Variance: It can be found that the variance of daily rise and fall of dog currency is the largest, 27.961, so the price of this currency fluctuates the most; The variance of Teda, USD Coin and Binance USD is 0, indicating that the price fluctuation of these three currencies is very small and the currency value is stable.

Maximum: The size of the maximum reflects the upper and lower limits of the daily rise and fall of each digital cryptocurrency in the statistical range, which is related to the mean value. In general, the larger the range reflected by the maximum, the larger the corresponding variance, but not absolute.

3.2 Interpretation of Cluster Analysis Results

Figure 2 shows the process of clustering. In the following steps, for a new category containing multiple samples, in fact, one sample of the class is used to represent the category; At the same time, we can clearly see the category of each digital cryptocurrency and observe the whole clustering process and results. This figure provides all the classification results of 1~10 categories. In this paper, it is divided into 3 categories. The types of digital cryptocurrencies included in each category are shown in the Table 1.

Table 1. Cluster analysis results

Category	Digital cryptocurrency	Quantity
First set	Bitcoin, Ethereum, BNB, Solana, Ida, XRP, Dog Coin	7
Second set	Teda	1
Third set	USD Coin, Binance USD	2

3.3 Explanation of the Result of Univariate Linear Regression

Figure 3 is a scatter diagram, which is plotted for DEPENDENT and * ZPRED. The two variables show a linear trend, indicating that the prediction effect of the equation is very good. The regression equation is obtained as follows:

$$y = 0.119 + 1.001x + \varepsilon \tag{7}$$

It can be seen that the price of Ethereum plummeted with the price of Bitcoin, not by accident, but because there is a significant correlation between them.

4 Conclusions

There is a certain correlation between various digital cryptocurrencies; The ten selected digital cryptocurrencies can be divided into three categories according to the method of system cluster analysis, and there is significant positive correlation between two pairs in each category; There is a strong positive correlation between Bitcoin and Ethereum, and the specific regression model can be obtained through the univariate linear regression analysis; The deep reason why the price of Ethereum plummeted with the price of Bitcoin is that there is a significant correlation between the two.

References

[1] B. Lan, L. Zhuang. (2021) Research on the Impact of the COVID-19 on the Financial Market. Statistics and Decision Making, 2021,37 (05): 129-133. DOI: 10.13546/j.cnki.tjyjc.2021.05.028.

[2] S. Song, D. Liu. (2021) Global competition of central bank digital currency: drivers, trade-offs and development trends. Comparison of economic and social systems, 2021, No.214 (02): 1-11.

[3] Y. Yuan, F. Wang. (2016) Development status and prospect of blockchain technology. Journal of Automation, 42 (04): 481-494. DOI: 10.16383/j.aas.2016.c 160158.

[4] J. Jiao, T. Sun, T. Huang. (2015) Digital currency and inclusive financial development - theoretical framework, international practice and regulatory system. Financial Regulation Research, No.43 (07): 19-35. DOI: 10.13490/j.cnki.frr.2015.07.002.
[5] T. Pi, C. You. (2022) Research on the dynamic correlation between cryptocurrency and traditional financial asset market -- empirical analysis based on t-Copula-DCC-GARCH. Financial Forum, 27 (05): 42-51. DOI: 10.16529/j.cnki.11-4613/f.2022.05.003.
[6] W. Zhou, Y. Li, Tan Jing. (2021) Research on the correlation between cryptocurrency and stock market risk. China Soft Science, 2021 (S1): 116-126.
[7] J. Sun, J. Liu, L. Zhao. (2008) Research on clustering algorithm. Journal of Software, 2008 (01): 48-61.
[8] Q. Zhou, Y. Liu, D. Wang. (2022) Evolution model of enterprise technology standard cooperation network oriented to the "the Belt and Road" - cluster analysis of social networks based on time series changes of countries. Science of Science and Management of Science and Technology, 2022,43 (09): 54-70.
[9] T. Lu, B. Tao, S. Zhou. (2008) Linear regression modeling and solution based on global least squares method. Journal of Wuhan University (Information Science Edition), 2008 (05): 504-507.
[10] X. Xiao, X. Wu. (2021) Geometric interpretation of multicollinearity in linear regression. Statistics and Decision, 2021,37 (21): 46-51. DOI: 10.13546/j.cnki.tjyjc.2021.21.009.