Robust Statistical Pearson Correlation Diagnostics for Bitcoin Exchange Rate with Trading Volume: An Analysis of High Frequency Data in High Volatility Environment

Nashirah Abu Bakar¹, Sofian Rosbi²

¹Islamic Business School, College of Business, Universiti Utara Malaysia, Malaysia
²School of Mechatronic Engineering, Universiti Malaysia Perlis, Malaysia

Abstract—Cryptocurrency is a digital or virtual currency that uses cryptography for security, transfer process and storage in ledger. This paper is to validate the correlation between exchange rate changes and trading volume changes. Data selected for this study is hourly data starting from 4 November 2017 until 7 November 2017. Methodology implemented in this study started with normality diagnostics and followed by correlation diagnostic. In this study, Pearson correlation calculation is implemented to evaluate the association between two variables namely exchange rate and trading volume. Pearson's correlation coefficient (r) is a measure of the strength of the association between the two variables. Result shows the coefficient of association is 0.123. Therefore, this study proved that the association between exchange rate changes and trading volume changes is very weak association. This value occurred because there is high volatility in hourly data and existence of outliers. The significant of this finding will help investors to recognize the relationship between trading volume and exchange rate. Therefore, it will help investors to make better decision in developing investment portfolio.

Keywords—Bitcoin, Volatility, Correlation, Exchange rate, Trading volume.

I. INTRODUCTION

Bitcoin cryptocurrency is defined as a digital currency in which encryption techniques are used to regulate the generation of units of currency. Bitcoin cryptocurrency involve with a complex process with the bitcoin cryptocurrency in which encryption techniques are used to regulate the generation of units of currency. Then the system will verify the transfer of funds, operating independently from central bank. It is, however, not subject to regulation by central banks, does not enjoy the backing of goods or services with intrinsic value (Rees, 2014).

Abu Bakar et al. (2017) explains the process of bitcoin cryptocurrency transaction procedure was started with the User A transfer digital currency to User B. The transaction needs to go through the blockchain path. A blockchain is an open, distributed ledger that can record transactions between two parties efficiently and in a verifiable and permanent way (Reid and Harrigan, 2013). A transaction is a transfer of bitcoin value that is broadcast to the network and collected into blocks. A transaction typically references previous transaction outputs as new transaction inputs and dedicates all input bitcoin values to new outputs (Miers, et al., 2013). Ledger is open to all users in the networks, and all users refer to one public ledger of transaction chain (Moore and Christin, 2013). Cryptocurrency make it easier to transfer funds between two parties in a transaction; these transfers are facilitated using public and private keys for security purposes. These fund transfers are done with minimal processing fees, allowing users to avoid the steep fees charged by most banks and financial institutions for wire transfers. Cryptocurrency defines an electronic coin as a chain of digital signatures (Okamoto, 1995). Each owner transfers the coin to the next by digitally signing a hash of the previous transaction and the public key of the next owner and adding these to the end of the coin. A payee can verify the signatures to verify the chain of ownership. Bitcoin miners help keep the Bitcoin network secure by approving transactions (Kroll et al., 2013). Mining is an important and integral part of Bitcoin that ensures fairness while keeping the Bitcoin network stable, safe and secure (Ron and Shamir, 2013).

Innovation in cryptocurrency was increasing popularity and the Bitcoin are expected to be a medium of exchange between the buyer and seller. Therefore, this study was performed to investigate the bitcoin transaction. Specifically, this study will validate the correlation
between exchange rate changes and trading volume changes.

II. LITERATURE REVIEW

Bitcoin cryptocurrency is different from conventional currency because it is not a fiat money or specific money. Bitcoin also not regarded as legal tender by a central authority or backed by goods or services having an intrinsic value (Christopher, 2014) and bitcoin is also decentralised in the sense that it is not issued by a government or single institution (Ram, et al, 2016). The price of bitcoin is based on supply and demand. The exchange rate of cryptocurrency fluctuate widely depend on news or speculations (Abu Bakar and Rosbi, 2017). According to Abu Bakar and Rosbi (2017), bitcoin cryptocurrency involved with high volatility. They found the standard error for Bitcoin volatility is 4.458 % show as high value of volatility.

A defining feature of a cryptocurrency and arguably its most endearing allure is its organic nature; it is not issued by any central authority, rendering it theoretically immune to government interference or manipulation (Bohme, et al., 2015). Most cryptocurrencies are designed to gradually decrease the production of currency, placing an ultimate cap on the total amount of currency that will ever be in circulation, mimicking precious metals (Barber, et al., 2012).

Abu Bakar and Rosbi (2017) conclude that a bitcoin transaction is a transfer of bitcoin value that is broadcast to the network and collected into blocks. A transaction typically references previous transaction outputs as new transaction inputs and dedicates all input bitcoin values to new outputs. Transactions are not encrypted, so it is possible to browse and view every transaction ever collected into a block. Once transactions are buried under enough confirmations, they can be considered irreversible. This system is vulnerable to hacking activity. Bitcoin cryptocurrency also has no physical form and exists only in a network. Bitcoin cryptocurrency is no intrinsic value in that it is not redeemable for another commodity, namely gold.

III. RESEARCH METHODOLOGY

This section describes the methodology implemented in this study starting from data selection, data transformation, normality diagnostics and Pearson correlation diagnostics. This study performed Pearson correlation analysis for exchange rate changes to trading volume changes.

3.1 Data selection

There are two variables are selected in this study namely exchange rate (USD/Bitcoin) and total trading volume (USD). Both of the variables is collected hourly starting from 4 November 2017 until 7 November 2017. These data are collected from https://www.worldcoinindex.com.

3.2 Mathematical derivation for exchange rate changes and trading volume changes

This study evaluates the correlation between exchange rate changes and trading volume changes. Therefore, calculation for exchange rate and trading volume changes need to be derived.

Firstly, this study derived the percentage changes for exchange rate using Equation (1).

\[ \Delta EX = \frac{EX_t - EX_{t-1}}{EX_{t-1}} \times 100\% \] .......................... (1)

Where:

- \( \Delta EX \) is percentage changes of exchange rate,
- \( EX_t \) is exchange rate value for trading period \( t \) and
- \( EX_{t-1} \) is exchange rate value for trading period \( t-1 \).

Next, this study derived the percentage of changes for trading volume as stated in Equation (2).

\[ \Delta TV = \frac{TV_t - TV_{t-1}}{TV_{t-1}} \times 100\% \] .......................... (2)

where:

- \( \Delta TV \) is percentage changes of exchange rate,
- \( TV_t \) is exchange rate value for trading period \( t \) and
- \( TV_{t-1} \) is exchange rate value for trading period \( t-1 \).

3.3 Normality statistical test

The probability density of the normal distribution is:

\[ f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \] .......................... (3)

Where:

- \( \mu \) is the mean or expectation of the distribution,
- \( \sigma \) is standard deviation, and
- \( \sigma^2 \) is variance for data distribution.

Properties of a normal distribution:

(a) The mean, mode and median are all equal.
(b) The curve is symmetric at the center. Data are distributed around the mean, \( \mu \).
(c) Exactly half of the values are to the left of center and exactly half the values are to the right.
(d) The total area under the curve is 1.

The null-hypothesis of Shapiro-Wilk normality test is that the population is normally distributed. Thus, if the p-value is less than the chosen alpha level, then the null hypothesis is rejected and there is evidence that the data tested are not from a normally distributed population. On the opposite side, if the p-value is greater than the chosen alpha level, then the null hypothesis that the data came from a normally distributed population cannot be rejected.
The Shapiro-Wilk test is a method to evaluate whether a random sample comes from a normal distribution. The test gives you a W value. The W value larger than chosen alpha (0.05), will concludes the distribution of data follows normal distribution. The, if the data shows small values of W, it is indicate your sample is not normally distributed. The formula for the W value is:

\[ W = \frac{\left( \sum_{i=1}^{n} a_i x_{(i)} \right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \]

where:
- \( x_i \) is the value in the sample \( (x_1, x_2, x_3, ..., x_n) \);
- \( x_{(i)} \) is the ordered sample values \( (x_{(i)}) \) is the smallest value in the sample);
- \( \bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} \) is the sample mean;
- \( a_i \) is constants that derived generated from the means, variances and covariances of the order statistics of a sample of size \( n \) from a normal distribution. The calculation of \( a_i \) is described in below equation.

\[ (a_1, a_2, a_3, ... , a_n) = \frac{m! V^{-1}}{(m! V^{-1} m)^{1/2}} \]

where:
- \( V \) is the covariance matrix of those order statistics;
- \( m = (m_1, m_2, m_3, ..., m_r)^T \)

Element in Equation (8) is represented as:

\[ m_1, m_2, m_3, ..., m_r \]

are the expected values of the order statistics of independent and identically distributed random variables sampled from the standard normal distribution.

### 3.4 Pearson correlation diagnostics

The Pearson product-moment correlation coefficient is a measure of the strength of a linear association between two variables and is denoted by \( r \). The Pearson product-moment correlation develops a line of best fit through the data of two variables. Then, the Pearson correlation coefficient, \( r \), indicates how well the data points fit this modeling line.

Consider the Pearson product-moment correlation coefficient of two \( n \)-dimensional vectors \( X = [X_1, X_2, ..., X_n] \) and \( Y = [Y_1, Y_2, ..., Y_n] \). Pearson correlation is states as the ratio between the covariance of \( X \) and \( Y \) and the product of their standard deviations. Pearson's correlation coefficient when applied to a population is commonly represented by below equation:

\[ \rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \]

where \( \text{cov} \) is the covariance, \( \sigma_X \) is the standard deviation of \( X \), and \( \sigma_Y \) is the standard deviation of \( Y \).

Then, covariance expressed as below:

\[ \text{cov}(X,Y) = E[(X - \mu_X)(Y - \mu_Y)] \]

where \( E \) is the expectation and \( \mu_X \) is the mean of \( X \).

Therefore, Equation (12) can be written as:

\[ \rho_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \]

Then, mathematical equation for \( \rho \) can be expressed in terms of uncentered moments. Mean of population is expressed as next equation,

\[ \mu_X = E[X], \mu_Y = E[Y] \]

Variance of population is expressed as next equation,

\[ \sigma_X^2 = E[(X - E[X])^2] = E[X^2 - 2XE[X] + (E[X])^2] \]

\[ = E[X^2] - 2XE[X] + (E[X])^2 \]

\[ = E[X^2] - 2E[X]^2 + (E[X])^2 \]

\[ = E[X^2] - E[X]^2 \]

\[ \sigma_Y^2 = E[Y^2] - E[Y]^2 \]

Standard deviation of population is expressed as next equation,

\[ \sigma_X = \sqrt{E[X^2] - E[X]^2}, \]

\[ \sigma_Y = \sqrt{E[Y^2] - E[Y]^2} \]

Covariance of population is expressed as next equation,

\[ E[(X - \mu_X)(Y - \mu_Y)] \]

\[ = E[(X - E[X])(Y - E[Y])] \]


\[ = E[XY] - E[X]E[Y] \]

Therefore, Equation (14) can be represented as:

\[ \rho_{X,Y} = \frac{E[XY] - E[X]E[Y]}{\sqrt{E[X^2] - E[X]^2} \sqrt{E[Y^2] - E[Y]^2}} \]

Then, the equation for sample is derived. Sample Pearson’s correlation coefficient is commonly represented by the letter \( r \). Consider the sample of dataset \( x = \{x_1, x_2, ..., x_n\} \) containing \( n \) values and another dataset \( y = \{y_1, y_2, ..., y_n\} \) containing \( n \) values then that formula for \( r \) is:

\[ r_{xy} = \frac{\text{cov}(x,y)}{s_x s_y} \]
where $\text{cov}$ is the covariance, $s_x$ is the standard deviation of $x$, and $s_y$ is the standard deviation of $y$.

Then, sample covariance can be expressed as below:

\[
\text{sample cov}(x, y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1}
\]

Therefore, Equation (16) can be written as:

Then, mathematical equation for $r$ can be expressed in terms of uncentered moments. Mean of sample,

\[
\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \quad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}
\]

Variance of sample,

\[
s_x^2 = \frac{1}{n} \left[ \sum_{i=1}^{n} (x_i - \bar{x})^2 \right] = \frac{1}{n} \left[ \sum_{i=1}^{n} \left( x_i - \bar{x} \right)^2 - 2 \bar{x} \sum_{i=1}^{n} x_i + n \bar{x}^2 \right]
\]

\[
= \frac{1}{n-1} \left[ \sum_{i=1}^{n} (x_i - \bar{x})^2 - n \bar{x} \bar{y} \right]
\]

\[
= \frac{1}{n-1} \left[ \sum_{i=1}^{n} (x_i - \bar{x})^2 - n \left( \frac{\sum_{i=1}^{n} x_i}{n} \right)^2 \right]
\]

\[
= \frac{1}{n-1} \left[ \sum_{i=1}^{n} x_i^2 - \left( \frac{\sum_{i=1}^{n} x_i}{n} \right)^2 \right]
\]

\[
= \frac{1}{n(n-1)} \left[ n \sum_{i=1}^{n} x_i^2 - \left( \frac{\sum_{i=1}^{n} x_i}{n} \right)^2 \right]
\]

\[
s_y^2 = \frac{1}{n} \left[ \sum_{i=1}^{n} (y_i - \bar{y})^2 \right] = \frac{1}{n-1} \left[ \sum_{i=1}^{n} y_i^2 - \left( \frac{\sum_{i=1}^{n} y_i}{n} \right)^2 \right]
\]

Standard deviation of sample,

\[
s_x = \sqrt{\frac{1}{n-1} \left[ n \sum_{i=1}^{n} x_i^2 - \left( \frac{\sum_{i=1}^{n} x_i}{n} \right)^2 \right]}
\]

\[
s_y = \sqrt{\frac{1}{n(n-1)} \left[ n \sum_{i=1}^{n} y_i^2 - \left( \frac{\sum_{i=1}^{n} y_i}{n} \right)^2 \right]}
\]

Covariance of sample,

\[
\text{cov}(x, y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1}
\]

\[
= \frac{1}{n-1} \left[ \sum_{i=1}^{n} x_i y_i - \bar{x} \bar{y} \right]
\]

\[
= \frac{1}{n-1} \left[ \sum_{i=1}^{n} x_i y_i - \left( \frac{\sum_{i=1}^{n} x_i}{n} \right) \left( \frac{\sum_{i=1}^{n} y_i}{n} \right) \right]
\]

\[
= \frac{1}{n-1} \left[ \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{j=1}^{n} y_j \right]
\]

\[
r_{x,y} = \frac{\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{j=1}^{n} y_j}{\sqrt{\left( \sum_{i=1}^{n} x_i^2 - \left( \frac{\sum_{i=1}^{n} x_i}{n} \right)^2 \right) \left( \sum_{i=1}^{n} y_i^2 - \left( \frac{\sum_{i=1}^{n} y_i}{n} \right)^2 \right)}}
\]

\[
IV. \quad \text{RESULT AND DISCUSSIONS}
\]

This section describes the result for statistical test of normality data characteristics. Then, this study performs Pearson correlation diagnostics to evaluate the association between changes of exchange rate with changes of trading volume.

3.1 Normality characteristics of data for exchange rate

This section describes the normality checking for data distribution of exchange rate. The function of this analysis is to validate the normality characteristics. Figure 1 shows the dynamic behavior of exchange rate. The value of exchange rate is referring as value of United States Dollar (USD) to the value for each of Bitcoin. Data were collected hourly starting from 4 November 2017, 00:00 until 7 November, 24:00. There are 97 observations. The maximum value is 7,557.82 USD on 6 November 2017.
03:00. Meanwhile, the minimum value is 6,996.70 USD on 7 November 2017, 16:00.

Then, this study calculated the percentages of changes with respect to previous observation period. Figure 2 shows percentage of changes for exchange rate of Bitcoin.

The maximum value is 2.3168 on 4 November 2017, 10:00. Mean of the data distribution is -0.0301 and standard deviation is 0.79764.

Next, this study validates the normality characteristics finding using histogram, normal probability plot and statistical test. Figure 3 shows the histogram of exchange rate changes. The distribution of exchange rate changes follows the normal distribution line. Figure 4 is normal probability plot of exchange rate changes in percentages. Data distribution is near to normal distribution line. Therefore, the data distribution is follow normal distribution.

Then, this study validated the normality using Shapiro-Wilk normality test. Table 1 shows the Shapiro-Wilk normality statistical test. The p-value is 0.567. Therefore, the data distribution is follow normal distribution.

3.2 Normality characteristics of trading volume data

This section describes the normality checking for trading volume data. Data selected in this study involving data of trading volume starting from 4 November 2017, 0:00 until 7 November 2017, 24:00. The minimum value of trading volume is 1.860x10^9 USD on 5 November 2017, 18:00. Meanwhile, the maximum value of trading volume is 2.998 x10^9 USD on 4 November 2017, 10:00.

Then, this study calculated the percentage of changes in the trading volume. Figure 6 shows the percentage of trading volume changes. The analysis shows the maximum value is 9.8619 on 6 November 2017, 13:00. Meanwhile, the minimum value is -6.4298 on 4 November 2017, 17:00. Mean of the data distribution for changes of trading volume is -0.38392. In addition, the standard deviation of data distribution for trading volume changes is 2.82637.
Next, this study performed normality test diagnostics to evaluate the data distribution of changes for trading volume data. This study implemented graphical approach and numerical statistical test approach to validate the normality of data distribution. Figure 7 shows the histogram for trading volume changes in percentage. The distribution is near to normal line. However, there are outliers in right side of normal distribution. Figure 8 shows the normal percentiles plot for trading volume changes. The distribution of data is near to reference line. Figure 8 indicates the present of outliers.

Table 2 shows the numerical prove of normality statistical test using Shapiro-Wilk approach. Probability value is 0.004. Therefore, data distribution is deviate from normal distribution. The presence of the outliers contributes to the non-normal distribution of data.

### Table 2: Normality test using Shapiro-Wilk

<table>
<thead>
<tr>
<th></th>
<th>Statistics</th>
<th>Degree of freedom</th>
<th>Probability value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trading volume changes</td>
<td>0.959</td>
<td>96</td>
<td>0.004</td>
</tr>
</tbody>
</table>

#### 3.3 Correlation diagnostics of exchange rate changes with trading volume changes

This section describes the correlation analysis between exchange rate changes with trading volume changes. Analysis that implemented in this section is using Pearson correlation method. First, this study validated the correlation using graphical method namely scatter plot between two variables. Figure 9 shows the scatterplot graph between trading volume changes and exchange rate changes. Next, this study validates the association between trading volume changes and exchange rate changes using Pearson correlation analysis. Table 2 shows the Pearson correlation diagnostics. Result shows the Pearson correlation is 0.123 that indicates very weak positive correlation. Significant value is larger than 0.05, this concludes there is no significant correlation between trading volume changes and exchange rate changes.
V. CONCLUSION
This objective of this study is to develop robust Pearson correlation diagnostics between trading volume changes and exchange rate changes. Data selected in this study are collected hourly starting from 4 November 2017, 0:00 until 7 November 2017, 24:00. Main findings of this study are described as below.

(a) In this study, two variables of data is collected namely exchange rate (USD/Bitcoin) and total trading volume (USD). Both of the variables are collected in hourly starting from 4 November 2017, 0:00 until 7 November, 24:00.

(b) This study calculated the percentages of exchange rate changes with respect to previous observation period. The maximum value is 2.3168 on 4 November 2017, 10:00. Mean of the data distribution is -0.0301 and standard deviation is 0.79764.

(c) This study validated the normality of exchange rate changes using Shapiro-Wilk normality test. Shapiro-Wilk normality statistical test shows the p-value is 0.567. Therefore, the data distribution is follow normal distribution.

(d) Then, this study calculated the percentage of changes in the trading volume. The analysis shows the mean of the data distribution for changes of trading volume is -0.38392. In addition, the standard deviation of data distribution for trading volume changes is 2.82637.

(e) Next, this study performed the numerical prove of normality statistical test using Shapiro-Wilk approach for trading volume changes. Probability value is 0.004. Therefore, data distribution is deviate from normal distribution. The presence of the outliers contributes to the non-normal distribution of data.

(f) Next, this study validates the association between trading volume changes and exchange rate changes using Pearson correlation analysis. Numerical result shows the Pearson correlation is 0.123 that indicates very weak positive correlation.

The findings of this study are important to investors and economics expert to validate the dynamic behavior of exchange rate associated with trading volume. High volatility environment contributes to the non-normality data distribution. In the same time, high frequency data for Bitcoin also indicates high volatility. Therefore, the finding of this study shows there is very weak positive correlation between trading volume changes and exchange rate changes.

REFERENCES
[5] Rees, M. (2014), Bitcoin to earth: don’t look now, but your paradigm is shifting”, *Bitcoin Magazine*, available at:


