The International Journal of Engineering and Science (IJES)
|| Volume || 7 || Issue || 4 Ver. II || Pages || PP 01-08 || 2018 ||
ISSN (e): 2319 – 1813 ISSN (p): 23-19 – 1805

Robust Outliers Detection Method For Ethereum Exchange Rate: A Statistical Approach Using High Frequency Data

Nashirah Abu Bakar, Sofian Rosbi
Islamic Business School, College of Business, Universiti Utara Malaysia, Malaysia
School of Mechatronic Engineering, Universiti Malaysia Perlis, Malaysia
Corresponding author: Nashirah Abu Bakar

--*ABSTRACT*--

Ethereum is one of the cryptocurrency that attracts attention from investors in year of 2017. Therefore, the objective of this study is to evaluate the data distribution of Ethereum exchange rate to validate the dynamic behavior of price movement. The finding of this study will help investors to understand the volatility of Ethereum exchange rate data. This study implemented Shapiro-Wilk normality test including graphical test to detect the outliers in the exchange rate data. The p-value of Shapiro-Wilk test is 0.0000. This value indicates the distribution of first difference for Ethereum exchange rate is not a normal distribution data. Then, this finding is validated using histogram and normal percentiles plot. Both of this plots indicates non-normal distribution because the data distribution does not follow normal distribution reference line. Finally, Box-Whisker plot is performed to detect the existence of outliers in the data. Result indicates there are suspected outliers and outliers in the Ethereum exchange rate data. This concluded that first difference of Ethereum exchange rate data is highly volatile. The important finding from this study is the dynamic behavior of Ethereum exchange rate is highly volatile and high risk. Therefore, any investors that interested with Ethereum cryptocurrency need to monitor closely the price to prevent high loss of their investment.

Keywords *– Cryptocurrency, Ethereum, Normality, Box-Whisker plot, Volatility.*

Date of Submission: 24-03-2018 Date of acceptance: 08-04-2018

I. INTRODUCTION

Cryptocurrency is a digital asset that was designed to work as a medium of exchange with decentralized monitoring system. In early 2014, Ethereum has received a lot of attention since its announcement at the North American Bitcoin Conference by Vitalik Buterin (Christoforou, 2017). The applications of Ethereum are run on its platform-specific cryptographic token known as ether (Browne, 2018). During 2014, Ethereum had launched a pre-sale for ether which had received an overwhelming response. Ether is like a vehicle for moving around on the Ethereum platform, and is sought by mostly developers looking to develop and run applications inside Ethereum. Ether is used broadly for two purposes; it is traded as a digital currency exchange like other cryptocurrencies and it is used inside Ethereum to run applications and even to monetize work (Browne, 2018).

Investing in cryptocurrency market is involved with high speculation and the market is largely unregulated by statutory bodies (Rees, 2014). Study by Abu Bakar and Rosbi (2017) regarding Bitcoin cryptocurrency found that the growth of Bitcoin exchange rate in moving towards non-equilibrium point. They conclude that the growth of the Bitcoin exchange rate is in unstable region. Even there are many study that focus on Bitcoin cryptocurrency, but there are still lack of researchers investigate the performance of Ethereum cryptocurrency transaction. Therefore this study try to fulfill this gap by examine the outlier of Ethereum cryptocurrency market. Outlier is important part in checking the normality of data distribution. Study from Abu Bakar and Rosbi (2017) regarding the outlier of data distribution for Bitcoin exchange rate found the data was follows non-normal distribution. Therefore, they implemented two-stages of outliers detection and deletion process and found the distribution of Bitcoin exchange rate with first difference is follow normal distribution with probability of 0.722.

II. LITERATURE REVIEW

Business through internet needs an electronic payment system that allowed two parties making a payment transaction. The common method of payment usually uses a credit card (Braunsberger et al., 2004; Foscht et al., 2010; Teoh et al., 2013). Currently, business through internet can be made using cryptocurrency that allowed two parties make a payment transaction using a digital currency. It also allowed two willing parties to transact directly with each other without the need for a trusted third party (Nakamoto, 2009; Reid and

Harrigan, 2013; Reynolds and Irwin, 2017). The innovation in financial technologies gives more advantage to financial institutions in making a payment transaction without using a fiat money. The ability of cryptocurrencies to enable anonymous transactions allows users to trade virtual currency regardless of their geographic location, without revealing either the real-world source of their income or their own identity (Reynolds and Irwin, 2017). Therefore, customers and sellers will get more advantage using an internet business.

Business through internet becomes a major form of communication, information dispersion tool and marketing tool in an organization (Razali, et al., 2010). As a marketing tool, internet gives wide coverage to reach potential client, better and quick information gathering and probably the cheapest way to sell the products. Internet also acts as a main component in e- commerce application. It will provide electronic services which company can use this application to sell their services (Razali, et al., 2010; Rahman, 2004). All this activities need customers to make a payment through online. Therefore, the existing of cryptocurrency give more advantage to organizations and customers.

A cryptocurrency transaction is a new tool in digital currency. Digital currency was allowed two willing parties to transact directly. This transaction are computationally impractical to reverse protect sellers from fraud, and routine escrow mechanisms could easily be implemented to protect buyers (Nakamoto, 2009). Blockchains are a software protocol that underlie cryptocurrency in one sense, are nothing more than a modernizing information technology, but in another sense, are novel and disruptive (Yeoh, 2017).

Cryptocurrency transaction is difference with traditional method of transaction. Figure 1 shows the cryptocurrency transaction procedure. If user A would like to transfer digital currency to user B, the transaction needs to go through the blockchain path. The blockchain in the ledger system that monitored and validated by the users in involved in ledger validation system using a computer system. Cryptocurrencies 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. There are no physical bitcoins; only balances kept on a public ledger in the cloud. All cryptocurrency transactions are verified by a massive amount of computing power (Abu Bakar et al., 2017). While study from Reynolds and Irwin (2017) explain the traditional conception of currency, that of tactile units or tokens of exchange, is termed as “real” currency, and includes the aspect of being supported and legitimized by each jurisdiction’s respective government, and is, thus, intrinsically tied to the restriction and introduction of supply by governments that affect its overall value within the market.

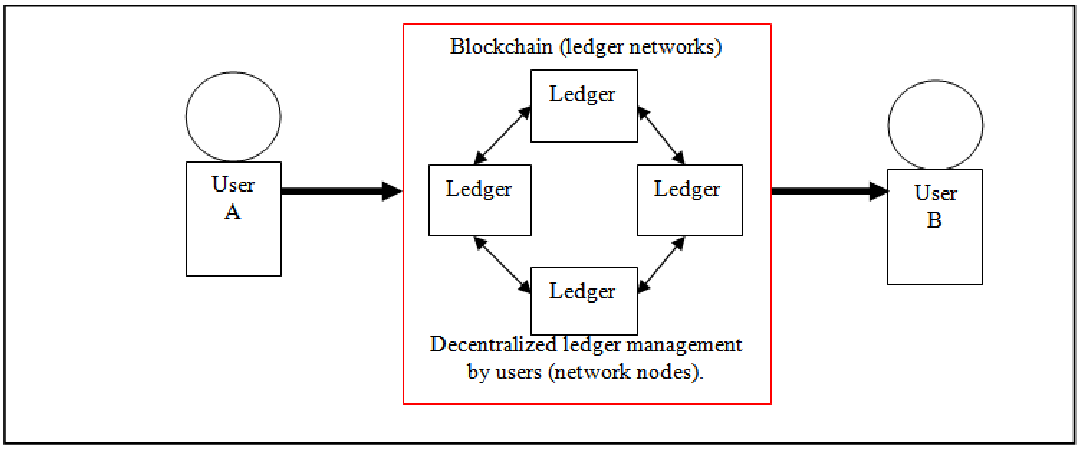

Figure 1: Cryptocurrency transaction procedure *(Sources: Abu Bakar et al. 2017)*

III. RESEARCH METHODOLOGY

The purpose of this study is to evaluate the normality of data distribution using Shapiro-Wilk statistical test. Next, this study implemented graphical statistical method to detect the present of outliers in data distribution.

3.1 Standard normal distribution

These studies evaluated the data distribution of Ethereum exchange rate using standard normal distribution. The equation for probability function is:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad \text{......} (1)$$

Equation 1 is based on assumption that mean, μ is zero and standard deviation, σ is 1. The coefficient $1/\sqrt{2\pi}$ is introduced to confirm the value of total area under the curve, $\varphi(x)$ is estimated equal to 1. Next, the exponential has factor of 1/2 to ensure variance and standard deviation is 1.The distribution function is symmetric around $x=0$. Next, the maximum value for probability function is $1/\sqrt{2\pi}$ at $x=0$. The inflection point is $x=+1$ and $x=-1$.

Figure 2 shows the normal distribution curve. Figure 2 describes 68.26 percentage of data are distributed between $\mu-\sigma$ and $\mu+\sigma$. Next, 95.44 percentage of data are distributed between $\mu-2\sigma$ and $\mu+2\sigma$. Then, 99.72 percentage of data are distributed between $\mu-3\sigma$ and $\mu+3\sigma$.

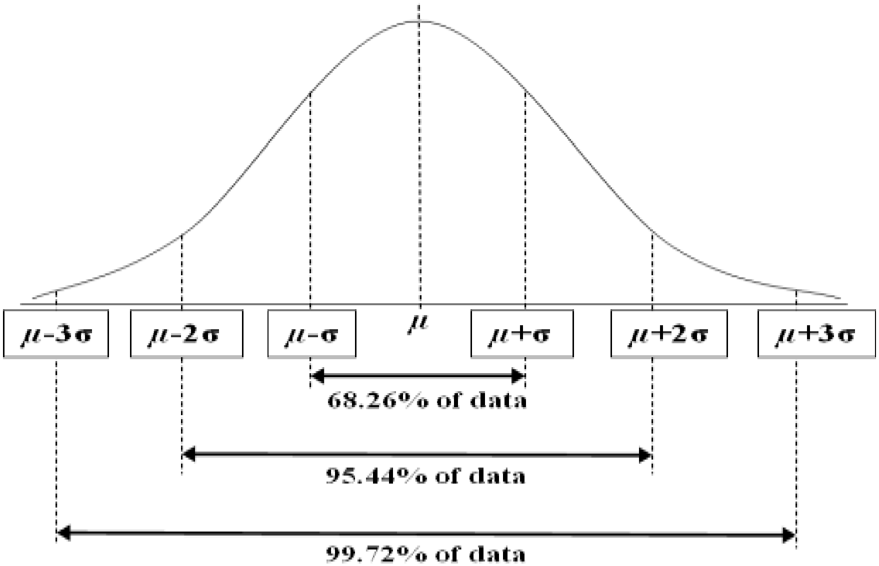

Figure 2: Standard normal distribution

The general equation for normal distribution of probability function can be expressed as:

$$f\left(x \middle| \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{......} (2)$$

In Equation (2), μ is mean of the data distribution, σ is standard deviation and σ^2 is the variance.

3.2 Derivation of mathematical equation for Shapiro-Wilk normality test

Shapiro-Wilk test is implemented to evaluate the normality of the data distribution. The null-hypothesis of this test is that the population is normally distributed.

Shapiro–Wilk considers a sample $x_1, x_2, x_3, ..., x_n$ is from a normal distribution population. The normality test statistics is represented by next equation:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \quad \text{......} (3)$$

In Equation (3), the parameter is defined as below:

(i) Variable $x_{(i)}$ is *i*-th order statistics.

The variable $x_{(1)}$ is first order statistics that indicates the minimum value of the sample.

$$x_{(1)} = \min\{x_1, ..., x_n\}$$

Meanwhile, variable $x_{(n)}$ is maximum order statistics for sample size of *n*.

$$x_{(n)} = \max\{x_1, ..., x_n\}$$

(ii) Sample is calculated using following equation of *n*, sample size:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n}$$

(iii) Then, the constant parameters a_i are calculated using below equation:

$$(a_1, a_2, a_3, ..., a_n) = \frac{m^T V^{-1}}{\left(m^T V^{-1} V^{-1} m\right)^{\frac{1}{2}}} \quad \text{.......} (4)$$

In Equation (4), m is the matrix for expected value of order statistics. The elements of matrix m ($m_1, m_2, m_3, ..., m_n$) are independent and identically distributed random variables from standard normal distribution. Matrix m is represented as below equation:

$$m = (m_1, m_2, m_3, ..., m_n)^T$$

Next, the symbol V is the covariance matrix of those order statistics.

(iv) The covariance matrix is derived as below procedure.

Consider matrix **X** is random vector. The elements of matrix **X** are random scalars with finite varience.

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ . \\ . \\ . \\ X_n \end{bmatrix}$$

In developing covariance equation for (i, j) dimension, there are two assumptions:

$$\mu_i = E(X_i), \mu_j = E(X_j)$$

Next, the covariance matrix is represented as below equation.

$$\begin{aligned} \textstyle\sum_{ij} = \operatorname{cov}(X_i, X_j) &= E\left[(X_i - \mu_i)(X_j - \mu_j)\right] = E\left[X_i X_j - \mu_j X_i - \mu_i X_j + \mu_i \mu_j\right] \\ &= E\left[X_i X_j\right] + E\left[-\mu_j X_i - \mu_i X_j + \mu_i \mu_j\right] \\ &= E\left[X_i X_j\right] + \left[-\mu_j \mu_i - \mu_i \mu_j + \mu_i \mu_j\right] \\ &= E\left[X_i X_j\right] + \mu_i \mu_j \end{aligned}$$

Therefore, the element for covariance matrix is described as below:

$$\sum = \begin{bmatrix} E\left[(X_1-\mu_1)(X_1-\mu_1)\right] & E\left[(X_1-\mu_1)(X_2-\mu_2)\right] & . & . & . & E\left[(X_1-\mu_1)(X_n-\mu_n)\right] \\ E\left[(X_2-\mu_2)(X_1-\mu_1)\right] & E\left[(X_2-\mu_2)(X_2-\mu_2)\right] & . & . & . & E\left[(X_2-\mu_2)(X_n-\mu_n)\right] \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ E\left[(X_n-\mu_n)(X_1-\mu_1)\right] & E\left[(X_n-\mu_n)(X_2-\mu_2)\right] & & & & E\left[(X_n-\mu_n)(X_n-\mu_n)\right] \end{bmatrix}$$

Next, the scalar value variance is described as below:

$$\sigma^2 = \operatorname{var}(X) = E\left[(X-E(X))^2\right] = E\left[(X-E(X))\cdot(X-E(X))\right]$$

Therefore, covariance matrix can be re-write as below:

$$\sum = \begin{bmatrix} \sigma^2 & E\left[(X_1-\mu_1)(X_2-\mu_2)\right] & . & . & . & E\left[(X_1-\mu_1)(X_n-\mu_n)\right] \\ E\left[(X_2-\mu_2)(X_1-\mu_1)\right] & \sigma^2 & . & . & . & E\left[(X_2-\mu_2)(X_n-\mu_n)\right] \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ E\left[(X_n-\mu_n)(X_1-\mu_1)\right] & E\left[(X_n-\mu_n)(X_2-\mu_2)\right] & & & & \sigma^2 \end{bmatrix}$$

IV. RESULT AND DISCUSSION

The objective of this study is to evaluate outliers existence in data distribution. Therefore, the methodology in this study started with data selection and first difference process. Next, this study performed numerical statistical test and graphical test to evaluate the outliers existence in Ethereum exchange rate data.

4.1 Data selection process

This study selected Ethereum exchange rate data from 19th January 2018, 00:00 until 31st January 2018, 23:55. The data is collected for every five minutes from *https://www.coindesk.com* . The total of observations is 3741. Figure 3 shows the dynamic behavior of Ethereum exchange rate data. The maximum value of the Ethereum exchange rate is on 28th January 2018, 12:10 (2739th observation). Meanwhile, the minimum value of Ethereum exchange rate is 23rd January 2018, 12:00 (1297th observation).

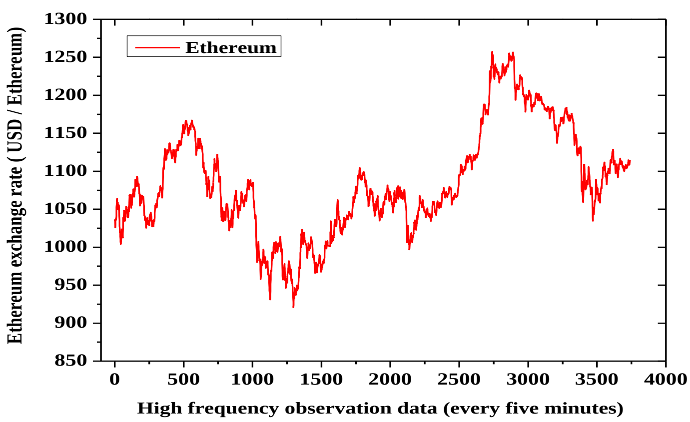

Figure 3: Dynamic movement of Ethereum exchange rate data

4.2 First difference process

Next, this study performed the first difference calculation to evaluate changes between current observation and previous observation. Figure 4 shows the first difference of Ethereum exchange rate data. The maximum value of first difference is 19 on 24th January 2018, 10:20 (1565th observation).

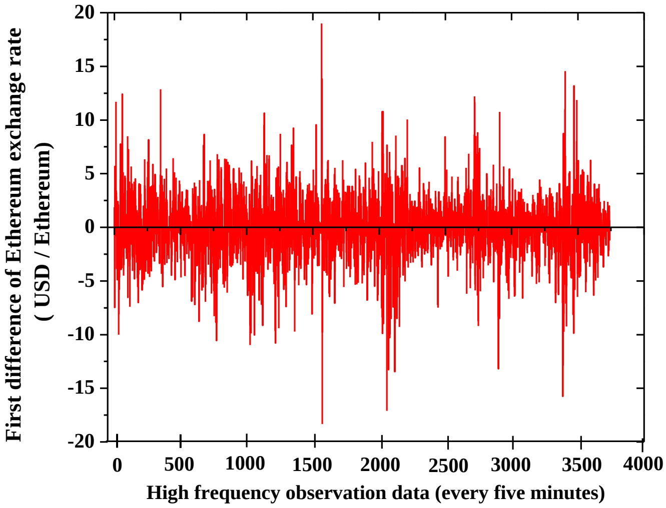

Figure 4: Rate of changes for Ethereum exchange rate

4.3 Numerical statistical test of normality

Next, this study performed normality statistical test to evaluate the data distribution for Ethereum exchange rate in first difference. Table 1 shows the result of Shapiro Wilk normality test for first difference of Ethereum exchange rate. The p-value is 0.0000 which less than 0.05, that is indicates this study reject null hypothesis. Therefore, the distribution of first difference of exchange rate is deviate from normal distribution.

Table 1: Shapiro-Wilk normality test

	Shapiro-Wilk normality test		
	Statistics	Degree of freedom	Significant (p-value)
First difference of exchange rate	0.965	3740	0.000

4.4 Graphical test for outliers detection

In this section, this study developed graphical method to analyze the non-normal distribution of data. Figure 5 shows the histogram of data distribution for first difference of exchange rate. Red line in Figure 5 shows the normal distribution standard as comparison to exchange rate distribution. On the both side of the tail distribution, there are observations that over the normal distribution line (red line). These data contribute to the non-normality characteristics for data distribution of Ethereum exchange rate.

Next, the normality probability is plotted to evaluate the existence of outliers in the data distribution. Figure 6 shows the normal probability plot for first difference of Ethereum exchange rate. Figure 6 shows there is data that is far from the normal distribution line (red line). Therefore, the existences of outliers contribute to non-normal data distribution.

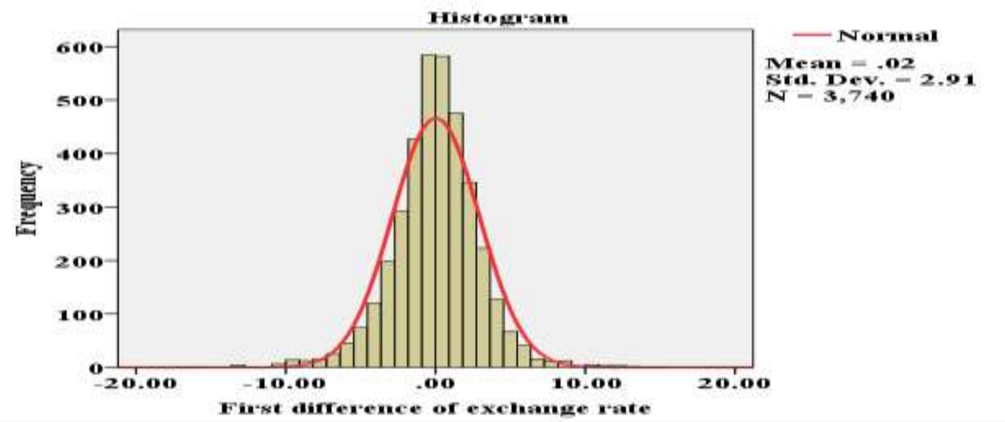

Figure 5: Histogram for first difference of exchange rate

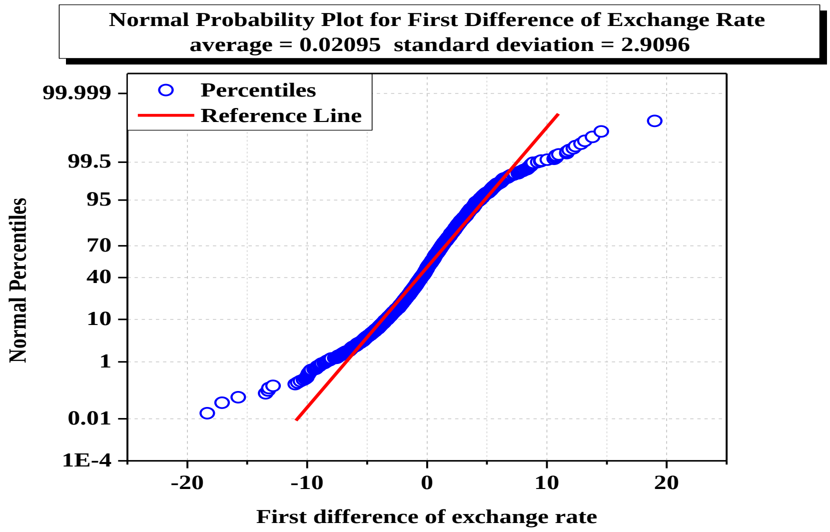

Figure 6: Normal probability plot for first difference of exchange rate

4.5 Outliers detection using Box-whisker plot

Finally, this study develop Box-Whisker plot to detect outliers in the data distribution. Figure 7 shows the Box-Whisker plot for first difference of Ethereum exchange rate. Figure 7 shows there are suspected outliers (blue circle mark) and outliers (red asterisks marks). Suspected outliers are data points that exist between the range of 1.5 and 3 inter quartile range (IQR). Meanwhile, outliers are data points that exist at the range over 3 IQR. Therefore, the existence of suspected outliers and outliers concluded that Ethreum exchange rata data is highly volatile.

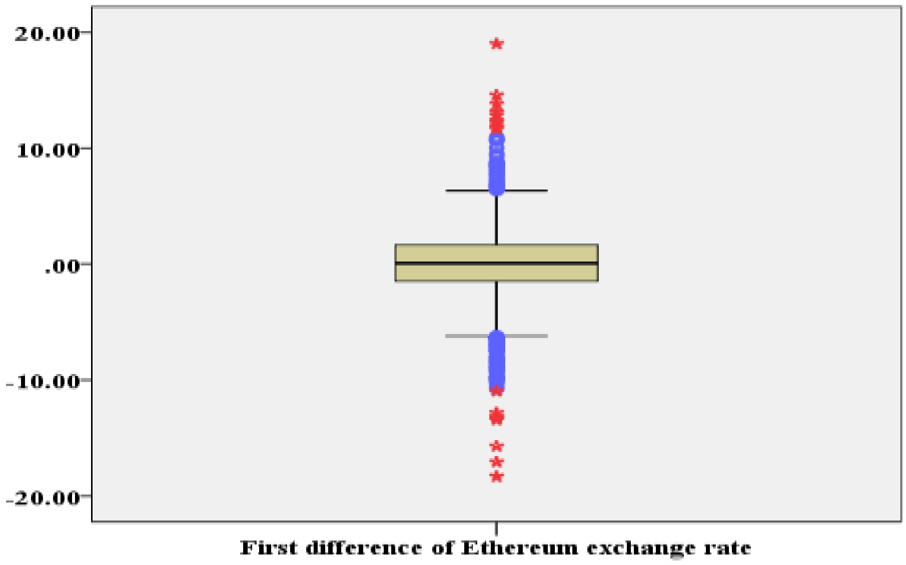

Figure 7: Box-Whisker plot for first difference of Ethereum exchange rate

V. CONCLUSION

The objective of this study is to detect outliers that contribute the volatility of Ethereum exchange rate data. This study recorded data starting from 19th January 2018, 00:00 until 31st January 2018, 23:55. The data is collected for every five minutes from *https://www.coindesk.com*. The total of observations is 3741. Main findings of this study are:

(a) The maximum value of the Ethereum exchange rate is on 28th January 2018, 12:10 (2739th observation). Meanwhile, the minimum value of Ethereum exchange rate is 23rd January 2018, 12:00 (1297th observation).

(b) Next, this study performed the first difference calculation to evaluate changes between current observation and previous observation. The maximum value of first difference is 19 on 24th January 2018, 10:20 (1565th observation).

(c) Next, this study performed normality statistical test to evaluate the data distribution for Ethereum exchange rate in first difference. The p-value is 0.0000 which less than 0.05, that is indicates this study reject null hypothesis. Therefore, the distribution of first difference of exchange rate is deviate from normal distribution.

(d) Then, this study developed graphical method to analyze the non-normal distribution of data. The histogram shows on the both side of the tail distribution, there are observations that over the normal distribution line. These data contribute to the non-normality characteristics for data distribution of Ethereum exchange rate.

(e) Next, the normality probability is plotted to evaluate the existence of outliers in the data distribution. Normality probability plot shows there is data that is far from the normal distribution line (red line). Therefore, the existences of outliers contribute to non-normal data distribution.

(f) Finally, this study develop Box-Whisker plot to detect outliers in the data distribution. The Box-Whisker plot shows there are suspected outliers and outliers. The suspected outliers are data points that exist between the range of 1.5 and 3 inter quartile range (IQR). Meanwhile, outliers are data points that exist at the range over 3 IQR. Therefore, the existence of suspected outliers and outliers concluded that Ethreum exchange rata data is highly volatile.

REFERENCES

[1]. A. Christoforou, Understanding the difference between Ethereum and Bitcoin, available at: http://theduran.com/understanding-the-difference-between-ethereum-and-bitcoin/, 2017.

[2]. F. Reid, and M. Harrigan, An Analysis of Anonymity in the Bitcoin System, Springer, New York, 2013.

[3]. K. Braunsberger, L.A. Lucas, and D. Roach, The effectiveness of credit-card regulation for vulnerable consumers, *Journal of Service Marketing*, 18(5), 2004, 358-370.

[4]. M. N. Razali, Z. Manaf and A. Mohd Yassin, Internet business strategies by property companies in Asia, *Property Management*, 28(4), 2010, 270-289.

[5]. M. Rees, Bitcoin to earth: don't look now, but your paradigm is shifting, Bitcoin Magazine, Retrieved from: http://bitcoinmagazine.com/15054/bitcoin-earth-dont-look-nowparadigm-shifting/ (accessed 25 March 2018), 2014.

[6]. N. Abu Bakar, S. Rosbi, and K. Uzaki, Cryptocurrency Framework Diagnostics from Islamic Finance Perspective: A New Insight of Bitcoin System Transaction, *International Journal of Management Science and Business Administration*, 4(1), 2017, 19-28.

[7]. N. Abu Bakar, and S. Rosbi, Identification of non-equilibrium growth for Bitcoin Exchange rate: Mathematical derivation method in Islamic Financial Engineering, *International Journal of Scientific Research and Management*, 5(12), 2017, 7772-7781.

[8]. N. Abu Bakar and S. Rosbi, Robust Statistical Normality Transformation method with Outlier Consideration in Bitcoin Exchange Rate Analysis, *International Journal of Advances in Scientific Research and Engineering*, 3(9), 2017, 80-91.

[9]. R. Browne (2018) Cryptocurrencies 'could drop to near-zero at any time,' Ethereum founder warns, Retrieved from: https://www.cnbc.com/2018/02/19/ethereum-creator-vitalik-buterin-warns-about-cryptocurrency-investment.html (accessed 25 March 2018) 2018.

[10]. P. Reynolds, and A.S.M. Irwin, Tracking digital footprints: anonymity within the bitcoin system, *Journal of Money Laundering Control*, 20(2), 2017, 172-189.

[11]. P. Yeoh, Regulatory issues in blockchain technology, *Journal of Financial Regulation and Compliance*, 25(2), 2017, 196-208.

[12]. S. Nakamoto, Bitcoin: A peer-to-peer electronic cash system", Retrieved from http://www.bitcoin.org /bitcoin.pdf (accessed 25 March 2018), 2009.

[13]. T. Foscht, C. Maloles, B. Swoboda, and S.L. Chia, Debit and credit card usage and satisfaction: who uses which and why – evidence from Austria, *International Journal of Bank Marketing*, 28(2), 2010, 150-165.

[14]. W.M.Y. Teoh, S.C. Chong, and S.M. Yong, Exploring the factors influencing credit card spending behavior among Malaysians, *International Journal of Bank Marketing*, 31(6), 2013, 481-500.

[15]. Z. Rahman (2004) E-commerce solutions for services, *European Business Review*, 16(6), 2004, 564-576.

Nashirah Abu Bakar ." Robust Outliers Detection Method For Ethereum Exchange Rate: A Statistical Approach Using High Frequency Data." The International Journal of Engineering and Science (IJES) 7.4 (2018): 01-08