JURNAL EMACS
(Engineering, MAthematics and Computer Science) Vol.7 No.1 January 2025: 107-114

e-ISSN: 2686-2573
DOI: 10.21512/emacsjournal.v7i1.12042

Research on the Empirical Analysis of Bitcoin and Gasoline Return

Asysta Amalia Pasaribu
Computer Science Department, School of Computer Science,
Bina Nusantara University,
Jakarta, Indonesia 11480
asysta.amalia@binus.ac.id

Correspondence: asysta.amalia@binus.ac.id

Abstract – *Investment is an activity that is popular nowadays. Profitable investments are the hope of every investor. By investing. investors expect the invested assets to generate returns and to obtain profits for future life In investment studies. the most frequently discussed topic is the fluctuations. whether increases or decreases. of an asset's price (stocks). The risk of investment is loss in financial. The fluctuations of stock prices represent risks in the investment field. One measure used to determine gains and losses from stock prices is the return. To know return from data. we may use the compound return formula. Returns have empirical facts that require several tests. In this study. the empirical facts of returns are that the returns are not autocorrelated (autocorrelation function) and that the returns are leptokurtic distributed (thick-tailed distribution). We use the price data of Bitcoin (BTC) and Gasoline (UGA) from January 1. 2019. to December 31. 2023. The main of purpose of this research is to show empirical analysis of the Bitcoin and Gasoline return data. The results of the empirical analysis show that the return of stock price for Bitcoin (BTC) and Gasoline (UGA) meet the empirical properties of returns so that they can capture a good volatility model.*

Keywords: Return, Autocorrelation Function (ACF), Bitcoin, Gasoline, Leptokurtic Distribution

I. INTRODUCTION

As people's incomes and needs increase today, people will look for ways to add income other than their main work output. People increasingly understand that to prepare for future needs is to invest. Investment is an activity of investing capital, directly or indirectly with the hope that in the future the owner of the capital will obtain certain benefits from the results of his investment. Investment is an event that contains risk. Risk is an event resulting from uncertainty in the future. The risk in investment activities is the return (Syuhada et all, 2021). Regarding risk, each investor has a different tolerance for investment risk. Someone who dares to take risks is called a risk-taker, someone who lacks courage or is hesitant (risk-moderate), while someone who does not dare to take risks is called a risk-averse.

People's goal in investing is to get a certain amount of profit in the future. The main concern for us is that the higher the expected return, the higher the risk that will be borne. This will make people understand the importance of returns in investment. In investing, returns have two numerical signs, where profit is a positive return. Losses are negative returns (Syuhada et al. 2021).

In the financial sector, types of investment are classified into 2 types, namely Real Assets and Financial Assets. Real assets are tangible assets, for example in the form of gold, land, property, patent rights and so on. Financial assets are intangible assets in the form of ownership documents that can be traded. Financial assets include money market instruments, shares, virtual currency, bonds and mutual funds. Return data from stock prices is an example of time series data. A return data is said to be good time series data that meets 2 conditions, namely the return data is not autocorrelated and the return data has a thick tail distribution. Therefore, the aim of this research is to analyze returns empirically using Bitcoin (BTC) and Gasoline (UGA) stock price data.

II. METHODS

2.1 The Definition of Return

Dangi (2023) stated that return functions to see the level of return from an investment over a certain time or period. Returns can be positive or negative. Positive return indicate profits earned for investors while negative returns indicate losses. The return formula consists of two calculations. namely simple return calculations and compound returns.

SIMPLE RETURN

Define P_t is the price of the asset at time t and P_{t-1} is the price of the asset at time $t-1$. Simple return (R_t) of an asset is defined as

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{1}$$

Furthermore, multi periode return can be calculated from time $t-k$ until time t.

$$\begin{aligned} 1 + R_t(k) &= \frac{P_t}{P_{t-k}} = \frac{P_t}{P_{t-1}} \cdot \frac{P_t}{P_{t-2}} \cdots \frac{P_{t-k+1}}{P_{t-k}} \\ &= (1+R_t)(1+R_{t-1}) \ldots (1 + R_{t-k+1}) \\ &= \prod_{j=0}^{k-1} (1 + R_{t-j}) \end{aligned} \tag{2}$$

Based on the multiplication sequence of multi-period returns in equation (2), it is difficult to calculate a return from the data, so that to determine the return value you can using the compound return formula.

COMPOUND RETURN

Define P_t is the price of the asset at time t and P_{t-1} is the price of the asset at time $t-1$. Compound return (R_t) of assets price has formed

$$R_t = ln \frac{P_t}{P_{t-1}} \tag{3}$$

Compound returns are related to as log-Return. Multi-period returns from time $t-k$ to time t can be calculated as follows.

$$\begin{aligned} R_t(k) &= \ln(1 + R_t(k)) \\ &= \ln[(1+R_t)(1+R_{t-1}) \ldots (1+R_{t-k+1})] \\ &= \ln(1+R_t) + \ln(1+R_{t-1}) \ldots + \ln(1+R_{t-k+1})] \\ &= R_t + R_{t-1} + R_{t-2} + R_{t-k+1} \end{aligned}$$

By using the definition of return, R_t, in equation (3), we can substitute into the following equation.

e-ISSN: 2686-2573
DOI: 10.21512/emacsjournal.v7i1.12042

$$= ln\frac{P_t}{P_{t-1}} + ln\frac{P_{t-1}}{P_{t-2}} + \cdots + ln\frac{P_{t-k+1}}{P_{t-k}}$$

$$= (\ln P_t - \ln P_{t-1}) + (\ln P_{t-1} - \ln P_{t-2}) + \cdots + (\ln P_{t-k+1} - \ln P_{t-k})$$

$$= (\ln P_t - \ln P_{t-1}) + (\ln P_{t-1} - \ln P_{t-2}) + \cdots + (\ln P_{t-k+1} - P_{t-k})$$

$$= (\ln P_t - ln\, P_{t-k})$$

To derive the compound return formula, simple return and compound return produce values that are not much different when viewed from the approximation using the Taylor series. The Taylor series has the following formula.

$$f(x) = f(c) + f'(c)(xc) + \frac{f"(c)(x-c)^2}{2!} + \cdots$$

Assume that function of $f(x) = \ln x$, then $\ln x = \ln c + \frac{(x-c)}{c} + \cdots$

If $x = 1 + R_t$, based on the simple return has the same value of $\frac{P_t}{P_{t-1}}$, so then:

$$\ln(1 + R_t) = ln1 + (1 + R_t - 1)$$

$$\ln(1 + R_t) = 0 + R_t$$

$$ln\left(\frac{P_t}{P_{t-1}}\right) = R_t$$

Calculating compound returns for multiple periods is easier to do by only paying attention to the difference in the ln function from the number of periods being calculated. This return calculation is called the additive nature of returns.

return periods is easier because it only looks at the nature of the difference ln functions from the specified number of periods. This is called the additive nature of returns. In this research. the return used is compound return.

2.2 The Empirical Analysis of Return

Return data is one of the example of time series data (observations). An observation series is said to be stationary if the process does not change along with changes in the time series.The test used for stationarity of time series data is the Augmented Dickey Fuller (ADF) stationary test. The purpose of the stationary test with Augmented Dickey Fuller (ADF) is stationary testing by determining whether the time series data contains a unit root. The ADF hypothesis test is as follows.

H_0 : Return data is not stationary

H_1: Return data is stationary

By using a significant test of α=0.05, decision making is carried out in two ways, namely, if the ADF statistical test > critical area test then H_0 is rejected. If the ADF statistical test < critical area test then H_0 is not rejected. Guo (2023) mentioned that the ADF test to test stationarity of financial time-series data. Nextmore, we will do empirical analysis for these return. Wang (2021) stated that there are empirical facts about returns, that returns has no autocorrelation, and returns have a leptokurtic distribution. The empirical properties of returns will fulfill and accommodate the volatility characteristics of returns.

RETURN HAS NO AUTOCORRELATION

The autocorrelation function (ACF) of a random variable is a measure of the correlation between two random variables in the same stochastic process (Syuhada et all, 2019). As the example. the return of R_t

is correlated with R_{t-k} at lag k denoted by ρ_k. Assume that R_t is weakly stationary. The autocorrelation coefficient with lag k is defined as:

$$\rho_k = \frac{Cov(R_t . R_{t-k})}{Var(R_t)} \tag{3}$$

This autocorrelation coefficient indicates the strength of the linear relationship between R_t with R_{t-k} which interval the value is in $-1 \leq \rho_k \leq 1$. Define R_t and R_{t-k} be random variables that are independent of each other, then $\rho_k = 0$ or it can be said that there is no linear relationship between of them.

The fluctuation of return in the investment has no significant autocorrelation. The autocorrelation of return will be trend rapidly. If return show significantly autocorrelation. then the autocorrelation can be used to develop a simple strategy to obtain positive profit expectations. In other words. we say that autocorrelation is expected to measure the relationship between a variable's present value and any past values that you may have access to.

The autocorrelation function (ACF) was carried out using the autocorrelation test (L-jung Box) with test statistics using the *chi-squared* distribution. The steps for the autocorrelation test with the Ljung-Box test are as follows.

1. $H_0 : \rho_1 = \rho_2 = \cdots = \rho_m = 0$
2. H_1 : There is $\rho_i \neq 0$ for $i \in 1. \ \ldots. m$
3. Significant level $\alpha = 0.05$
4. The test statistics used are as follows

$$Q(m) = T \sum_{k=1}^{m} \hat{\rho}_k^2$$

The $Q(m)$ function gets near to $chi-squared$ random variable with. m. the degrees of freedom

5. The critical area : $Q(m) > x_m^2$
6. H_0 is rejected in a critical area and H_0 is not rejected. if it is not in a critical area.

RETURN HAS LEPTOKURTIC DISTRIBUTION

Data in general follows a certain distribution. Kurtosis is a measure that can be used to determine a certain distribution of data. The measure of sharpness (kurtosis) is a measure used to determine the peak of a particular distribution following leptokurtic (taper), mesokurtic (normal), and platykurtic (blunt) (Untari, 2020). A graphic illustration of leptokurtic, mesokurtic and platykurtic distribution can be seen in **Figure 1.**

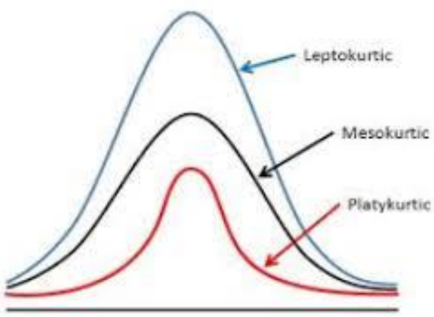

Figure 1. Leptokurtic, Mesokurtic, and Platykurtic

Based on Figure 1, it is explained that the shape of the height or sharpness of the peak of the curve is different, a distribution whose curve tends to be flat and the peak of the curve is not too high is platykurtic distribution, a curve whose peak is not too high or flat is mesokurtic distributin, and a distribution where the peak of the curve looks sharp and high is leptokurtic distribution.

Kurtosis has degrees of sharpness, namely Platikurtic distribution has a degree of kurtosis <3, Mesokurtic distribution with a degree of kurtosis = 3, and Leptokurtic distribution with a degree of kurtosis >3. Wang (2021) stated that the unconditional distribution of returns has sharp peaks and sharp tails tebal. The heavy tail distribution has a kurtosis with a value of 3 to 50. This indicates that the distribution of returns is extreme and not normal. The heavy tail also shows that the tail of the return distribution is slower towards zero when compared to the normal distribution.

e-ISSN: 2686-2573
DOI: 10.21512/emacsjournal.v7i1.12042

To determine the leptokurtic distribution of data. firstly we determine kurtosis using descriptive statistics. When kurtosis is more than 3. it can be said that the return distribution is not normally distributed. Another idea to find out that returns are not normally distributed is using a normality test with several methods, such as

1. Normal Probability Plot

Normal Probability Plot obtained using Rstudio software to determine the normal distribution of data. Normal Probability Plot represents a line connecting the 25th and 75th percentiles of the data. If the data points are along a straight line. then it can be said that the data is normally distributed. but conversely. if the data points are far from the line. then the data is not normally distributed.

2. Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov test is a test for the normality of a distribution. The steps for the normality test with Kolmogorov-Smirnov are as follows.

a) $H_0: F_n(x) = F^*(x) = \phi(x)$
b) $H_1: F_n(x) \neq F^*(x)$
c) The statistic test is

$$D = max_{t \le x \le u} |F_n(x) - F^*(x)|$$

with $F_n(x)$ is the cumulative distribution function of the data and $F^*(x)$ is the normal cumulative distribution function of data.

d) The Critical value : $\frac{1.22}{\sqrt{n}}$. $\frac{1.36}{\sqrt{n}}$. and $\frac{1.63}{\sqrt{n}}$ for confidence levels of 10%. 5%. and 1%.
e) If the calculation result D is smaller than the critical value or *p-value* greater than a certain level of confidence. Then. H_0 is not rejected. which means the data is normally distributed.

III. RESULTS AND DISCUSSION

In this paper we use daily the closing price data for Bitcoin (BTC) and Gasoline (UGA) from 01 January 2019 to 31 December 2023 which was downloaded on the website https://finance.yahoo.com/. The Bitcoin and Gasoline price displays in Figure 2.

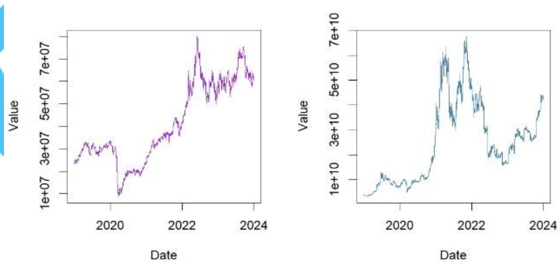

Figure 2. Gasoline Stock Price (left) and Bitcoin Stock Price (right)

In 2020, the fluctuation of Bitcoin (BTC) stock price saw a decline in prices. Meanwhile, Gasoline (UGA) stock prices start to increase in that time. The economic shut down due to the COVID-19 Pandemic. Bitcoin's price burst into action once again. According to The Forbes, the Covid-19 pandemic struck. and the stock markets dropped violently in mid-March 2020.

After 2020

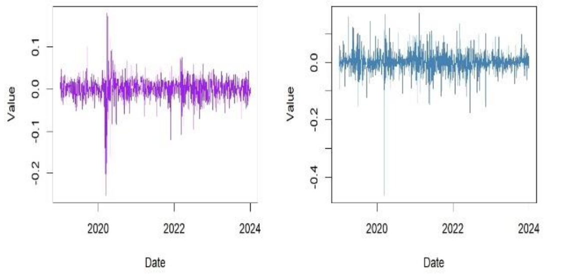

Figure 3. Return of Gasoline (left) and Bitcoin (right)

After calculating compound returns. the return data graph for Bitcoin and Gasoline is presented in Figure 2 belowBitcoin's there was a significant increase in the price of Gasoline (UGA) until 2022. After 2022. there was a decline in Bitcoin stock prices until the end of 2023. Based on the stock price data in Figure 2. the value of profits and losses in stock prices cannot be measured. Therefore. defining losses and profits from a stock investment. calculations can be made using the compound returns formula. return will decrease and increase in 2022. A significant decrease in Gasoline returns will occur in 2022. After 2022. Bitcoin and Gasoline returns will be stable.

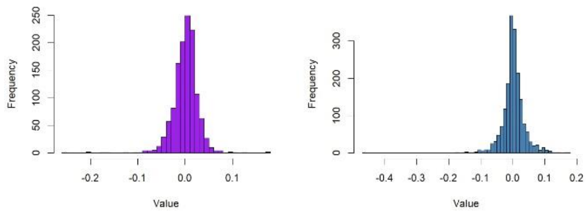

Figure 4. Histogram of Gasoline (left) and Bitcoin (right) Return

After calculating the compound return. Descriptive statistics for Bitcoin and Gasoline returns will be calculated which are presented in Table 1 below.

Table 1. Descriptive Statistics of Bitcoin and Gasoline Return

Statistic	BTC	UGA
n	1826	1258
Minimum	-0.4647	-0.2529
Maximum	0.1718	0.1795
Mean	0.0013	0.0008
Median	0.0008	0.0027
Variance	0.0012	0.0007
Skewness	1.2262	1.2699
Kurtosis	21.9658	17.0992

The number of observation (n) for Bitcoin Return and Gasoline Return respectively is 1826 and 1258. The minimum of return value for Bitcoin (BTC) is smaller than the return for Gasoline (UGA). The maximum return value for Gasoline is greater than the return for Bitcoin. The maximum value (profit) of Bitcoin is 0.1718 (17%). The maximum value (profit) of return for Gasoline is 0.1795. The Bitcoin data variance is smaller than the Gasoline data variance.

Skewness returns on Bitcoin (BTC) and Gasoline (UGA) are positive. so the distribution of both data has an elongated tail shape on the right or can be interpreted as meaning that the data tends to have low value. The kurtosis of Bitcoin and Gasoline return data is more than 3 degree. which means that the data has a leptokurtic distribution. Furthermore. stationary test will be carried out for Bitcoin (BTC) and Gasoline (UGA) return data using the ADF test using RStudio software. the *p-value* results of which are presented in the following Table 2.

Table 2. ADF Test Results of Bitcoin and Gasoline Return

	Bitcoin	Gasoline
p-value	0.01	0.01

The ADF test result of Bitcoin and Gasoline return in Table 2, which *p-value* is smaller than significant level $\alpha = 0.05$. Then, H_0 is rejected. So, Bitcoin and Gasoline return are stationary. Furthermore. an autocorrelation test will be carried out for Bitcoin (BTC) and Gasoline (UGA) return data using the Box-Pierce test using RStudio software. the *p-value* results of

e-ISSN: 2686-2573
DOI: 10.21512/emacsjournal.v7i1.12042

which are presented in the following Table 3.

Table 3. Autocorrelation Test Results of Bitcoin and Gasoline Return

	Bitcoin	Gasoline
p-value	0.9937	0.1976

By using a significance level of 0.05. the autocorrelation test results of Bitcoin and Gasoline returns show that the p-value is greater than 0.05. so H_0 is not rejected. This means that Bitcoin (BTC) and Gasoline (UGA) return data are not autocorrelated.

We can see in Figure 4 that Bitcoin and Gasoline return have not normal distribution. Futhermore, the result indicates that Bitcoin and Gasoline return has not mesokurtic distribution.

The distribution can be seen through the normal distribution plot presented in the following figure 4.

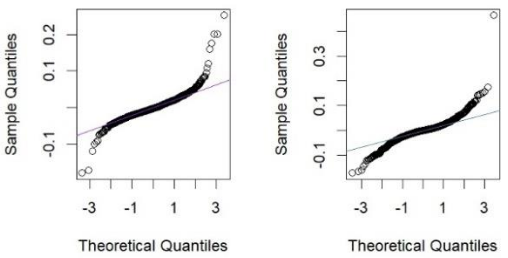

Figure 5. Normal Distribution Plot of Gasoline (left) and Bitcoin (right) return

Based on Figure 5, it is explained that the shape of the quantile points is located away from a straight line. so that the return distribution of Bitcoin (BTC) and Gasoline (UGA) is not normally distributed. Therefire, Bitcoin (BTC) and Gasoline (UGA) data normality test was carried out. The *p-value* of normality test results obtained in Table 2 are as follows.

Table 3. Normality Test of Bitcoin and Gasoline Return

	Bitcoin	Gasoline
n	1826	1258
p-value	0.0000	0.0000

The following Table 2 shows that *p*-value is smaller than confidence level. Therefore. Bitcoin (BTC) and Gasoline (UGA) return are not normal distribution.

IV. CONCLUSION

The gain return obtained from Gasoline are greater than Bitcoin which can be known through the Maximum value. Meanwhile. bigger losses occurred in Bitcoin. The autocorrelation test of Bitcoin and Gasoline returns shows that returns are not autocorrelated using the Ljung-Box test. This means that current Bitcoin and Gasoline returns have no relationship with the past return. Leptokurtic distribution in Bitcoin and Gasoline can obtained in Kurtosis. Normality Plot. and Kolmogorov Smirnov. Therefore. return is not normally distributed and has leptocurtic distribution. According to empiricsl analysis Bitcoin (BTC) and Gasoline (UGA) can capture the good volatility model. For further study. we can make a volatility model for Bitcoin and Gasoline return.

REFERENCES

Arum. P. R.. & Amri. S. (2024). Pemodelan Seasonal Autoregressive Integrated Moving Average (SARIMA) untuk Meramalkan Volume Angkutan Barang Kereta Api di Pulau Jawa

Tahun 2021: Memberikan prediksi volume angkutan barang kereta api di pulau jawa untuk masa mendatang. *Journal Of Data Insights*. *2*(1). 26-35.

Budianti. L.. Janatin. J.. Avicenna. M. Y.. Putri. A. K.. & Darmawan. G. (2024). Pemodelan SARIMA dengan Pendekatan ARCH/GARCH untuk Meramalkan Penjualan Ritel Barang Elektronik. Innovative: Journal Of Social Science Research. 4(1). 1037-1051.

Bondarenko. O.. & Muravyev. D. (2023). Market return around the clock: A puzzle. *Journal of Financial and Quantitative Analysis*. *58*(3). 939-967.

Dangi. V. (2023). An Econometric Study on Volatility Clusters. Dynamic Risk Return Relationship. and Asymmetry in Bitcoin Returns. Business Perspectives and Research. 22785337231208300.

Louarn. P.. Fedorov. A.. Prech. L.. Owen. C. J.. D'Amicis. R.. Bruno. R.. ... & Bale. S. D. (2024). Skewness and kurtosis of solar wind proton distribution functions: The normal inverse-Gaussian model and its implications. *Astronomy & Astrophysics*. *682*. A44.

Sihombing. E. I.. Suhendra. C. D.. & Marini. L. F. (2024). Analisis Data Time Series Untuk Prediksi Harga Komoditas Pangan Menggunakan Autoregressive Integrated Moving Average. KLIK: Kajian Ilmiah Informatika dan Komputer. 4(6). 2711-2720.

Suharmanto, B., & Ernawati, I. (2023, December). Uji Performa Prediksi Metode Autoregressive Fractionally Integrated Moving Averages Dan Long Short-Term Memory Dengan Data Saham Dua Perusahaan Bank. In *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya* (Vol. 4, No. 2, pp. 450-460).

Syuhada. K.. Wibisono. A.. Hakim. A.. & Addini. F. (2021). Covid-19 risk data during lockdown-like policy in Indonesia. *Data in Brief*. *35*. 106801.

Wang. C. (2021). Different GARCH models analysis of returns and volatility in Bitcoin. Data Science in Finance and Economics. 1(1). 37-59.

Weiß. C. H. (2024). The Mollified (Discrete) Uniform Distribution and its Applications. *arXiv preprint arXiv:2403.00383*.

Wicaksono, A., Farizan Gibran, A., Irmansyah, D., & Aji, H. (2021). Ukuran Penyebaran Data (Kemiringan Dan Keruncingan). Jurnal Ukuran Penyebaran Data (Kemiringan & Keruncingan), 1–6.

Zhichao, G. (2023). Research on the Augmented Dickey-Fuller Test for Predicting Stock Prices and Returns. In *Proceedings of the 7th International Conference on Economic Management and Green Development* (pp. 2754-1169).