

# Correlation Between Bitcoin Price and Total Supply of Long-term Holders

Tingfei Chen

North London Collegiate School (Singapore), 109708, Singapore

201273@student.nlcssingapore.sg

**Abstract.** This paper discusses the relationship between Bitcoin price and the number of Bitcoins held by Long-term holders (LTHs) through VAR and ARMA-GARCH models. The research and study continue to analyze the volatility of Bitcoin price while adding possible variables that modify the models better. The purpose is to better learn the market mechanics of Bitcoin and forecast price changes and long-term holders' behaviors. In this paper, two relationships are established: first, there's a long-term negative correlation between an increase in price and the number of Bitcoins held by LTHs. Second, the volatility of price and the volatility of the number of Bitcoins held by LTHs are positively correlated. Compared to previous analyses of the volatility of Bitcoin price, this paper considered a possible influential factor, the supply of LTHs, during the research. Based on the findings, suggestions for future investigations are given to consider short-term holders' relationship of the two variables mentioned in this paper. Financial institutions can better forecast the behaviors of other LTHs, whereas individual investors can have a better understanding of the market mechanism of Bitcoin while learning the behaviors of LTHs.

**Keywords:** Cryptocurrency, Bitcoin, Long-term Holder, ARMA-GARCH model, VAR model.

## 1. Introduction

From 2007 to 2008, the global economy had been largely raged by the Global Financial Crisis (GFC). Some of the reasons that caused GFC are risky housing loans, lack of regulation, and policy failure. As a result of the GFC, the desire for no government control over currency and untrustful financial institutions rose among the public [1]. Against this background, the ever-first cryptocurrency Bitcoin was created. Bitcoin is a digital form of currency that is decentralized, invented by a programmer or a group of programmers known as Satoshi Nakamoto in 2008.

Without the needs of governments and banks, the transaction of Bitcoins relies on a system called public-key cryptography. In this system, a private piece of data called the "private key" proves the user and only that user has access to a particular wallet. This mathematical mechanism provides Bitcoin buyers and sellers with a high-security level of protection. These protected transactions are recorded by using a structure called blockchain, which can be separated into block and chain. In each block. In every block, about 4000 transactions can be recorded. Connecting all blocks, the blockchain with public records of all transactions is produced. Since blockchain is shared with all users, it avoids the possibility of double spend [2]. Due to the transparency of bitcoins, the unspent transaction output (UTXO) system (i.e., amounts of digital currency that remained after the transaction) made it possible for users to track "coin ages", which is determined by the time of the last transaction.

Currently, because of the high volatility of the price of Bitcoins, researchers suggest it's more likely to be a speculative investment rather than real currency [3]. However, based on this property, past researchers chose to focus on investigating and explaining the return volatilities. In earlier studies, the GARCH models had been mainly applied for comparison between Bitcoins, gold, and US dollars [4]. However, past research mainly focused on analyzing and comparing the volatility of Bitcoin prices. There is a lack of analysis on factors of influence the price returns and volatilities.

One factor that has a strongly correlated pattern with the price of Bitcoin is the number of Bitcoins held by Long-term Holders (LTHs). Bitcoin holders are separated into two different categories based on different coin ages. The two categories are LTHs mentioned above and Short-term Holders (STHs).

The former is represented by those who are not sensitive to price changes, like financial institutions. The latter is represented by individual investors in the market.

In a previous analysis done by Rafael Schultze-Kraft [5], the coin age of the former has to be higher than 155 days, whereas that of the latter should be less than 155 days. This study separated LTH and STH by finding the gradient of the correlation between the probability of being spent in “x” days and coin age. The results show that under all cases (different x values) analyzed, the last date with the highest gradient (representing the last highest probability of being spent by an STH inside 1 year) always happens exactly at day 155. After holding for 155 days, the probability of the user being LTH, who is more inelastic to market price and thus less speculative, increases.

BTC: Long-Term Holder Supply, BTC: Price



**Figure 1** Total Supply Held by LTHs with Bitcoin Price

Data source: Glassnode[6]

Photo credit: Glassnode

As seen in Figure 1 above, there is a significantly strong relation between the logarithmic price (US dollar) of Bitcoin (hereinafter referred to as price) and the number (BTC) of Bitcoins held by LTHs(hereinafter referred to as supply). The turning points of supply are always associated with the situation inside the market, such as the bull and bear cycle.

Uncertainty of future prices has introduced two motivations, one of the motivations is that investors with large endowments will want to hedge by taking short positions, thus these highly risk-averse individuals (in the case of this paper are LTHs) will transfer risks of holding stake to speculators, the less risk-averse individuals (STHs) [7]. Another study shows when a market is in a volatile condition, individual investors (STHs) will have more challenges in terms of forecasting future prices. Information that is available for them is from earnings news, which conducts different beliefs. On the other hand, the trading volume of financial institutions and big firms (LTHs) that have more information will increase, showing how differences in holders will influence the change in trading volume [8]. Similarly, the correlation between price changes and trading volume has long been the topic of study. In the research of Jonathan M. Karpoff, the positive correlation between the magnitude of price change and trading volume was concluded [9]. Furthermore, another investigation on the link between price momentum and trading volume was conducted. The study shows a short-term(eight quarters) negatively correlated relationship between turnover ratio and future returns. However, this relationship will be reversed after 5 years, as higher trading volume leads to higher

returns. This correlation may be able to be applied to learn the market mechanism of Bitcoin price [10]. Moreover, the correlation was not a one-sided influence. Based on a study by Theoharry Grammatikos and Anthony Saunders, under most conditions, price change and trading volumes are contemporaneously correlated [11].

As a result, in this paper, the relation between Long-term Holders (LTHs) and the price of bitcoins will be investigated. Specifically, it focuses on the behavior of LTHs, and the number of bitcoins held by them. In the research, the VAR model is applied to test the dynamic relation between price and supply. Also, the ARMA-GARCH model is used to forecast future values by considering the conditional heteroskedasticity shown in both variables.

## 2. Research Design

### 2.1. Data Source

As the first cryptocurrency, Bitcoin has the largest amount of data with a longer timeframe than other cryptocurrencies. Therefore, more data are available for analysis. As discussed in 1, Bitcoin and other cryptocurrencies are unique compared to traditional currency. One of the uniqueness is that all transactions of Bitcoin are visible to the public. The paper extracts data from Glassnode, an intelligence platform for blockchain data. Data on the price (USD) and the number of BTC held by LTHs(supply) from 2010 July 17<sup>th</sup> to 2024 February 2<sup>nd</sup> are particularly focused. This data source will be used to study the correlations between the price of Bitcoin and the number of Bitcoins held by LTHs. Data processing is required to accomplish this goal. To get logarithmic price and logarithmic supply, the formula  $\ln(1 + x)$ , where x equals to price or supply, is used. Returns in price and supply are also required for further research. A return is produced by subtracting the prices or supply of the second day of two consecutive days by the price or supply of the first day, then dividing the results by the data of the first day. Last use the same formula  $\ln(1 + x)$ , where x equals the return of price or supply, to calculate the logarithmic price and supply return.

### 2.2. Augmented Dickey-Fuller (ADF) Test

Before starting model constrictions, testing the stationarity of the data should be tested. The null hypothesis for the ADF test, also known as the unit root test, is “the time series contains a unit root and is non-stationary”. In Table 1, the p-values of logarithmic price and supply return are all equal to 0, which is highly statistically significant. Therefore, rejecting the null hypothesis as they’re stationary and don’t contain a unit root.

**Table 1** ADF test

Variables	t-statistic	p-value
Raw		
Supply	-7.544	0.0000***
Price	-1.685	0.7575
Return		
Supply	-30.538	0.0000***
Price	-50.774	0.0000***

### 2.3. Construction of Vector Autoregression (VAR) Model

The Vector Autoregression (VAR) model is mainly used to predict multiple variables and their dynamic dependencies. In this bivariate VAR(p) model, time series price and supply are used as dependent variables in two equations, where the independent variable is the lag p variable of price and supply.

$$B_{1t} = \delta_1 + \alpha_{11}B_{1,t-1} + \dots + \alpha_{1p}B_{1,t-p} + \beta_{11}B_{2,t-1} + \dots + \beta_{1p}B_{2,t-p} + \varepsilon_{1t} \tag{1}$$

$$B_{2t} = B_2 + \alpha_{21}B_{1,t-1} + \dots + \alpha_{2p}B_{1,t-p} + \beta_{21}B_{2,t-1} + \dots + \beta_{2p}B_{2,t-p} + \varepsilon_{2t} \tag{2}$$

$$\begin{pmatrix} B_{1t} \\ B_{2t} \end{pmatrix} = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} + \begin{pmatrix} \alpha_{11} \\ \alpha_{21} \end{pmatrix} B_{1,t-1} + \dots + \begin{pmatrix} \alpha_{1p} \\ \alpha_{2p} \end{pmatrix} B_{1,t-p} + \begin{pmatrix} \beta_{1p} \\ \beta_{2p} \end{pmatrix} B_{2,t-1} + \dots + \begin{pmatrix} \beta_{1p} \\ \beta_{2p} \end{pmatrix} B_{2,t-p} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \quad (3)$$

$$\begin{pmatrix} B_{1t} \\ B_{2t} \end{pmatrix} = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \dots & \alpha_{1p} \\ \alpha_{21} & \dots & \alpha_{2p} \end{pmatrix} \begin{pmatrix} B_{1,t-1} \\ \vdots \\ B_{1,t-p} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \dots & \beta_{1p} \\ \beta_{21} & \dots & \beta_{2p} \end{pmatrix} \begin{pmatrix} B_{2,t-1} \\ \vdots \\ B_{2,t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \quad (4)$$

Equations (1) and (2) represent price and supply returns respectively. Because two equations have the same independent variable, putting (1) and (2) together equation (3) is formed, while (4) is (3) in matrix form. Specifically, in the equation of price return(1),  $\alpha_{10} + \alpha_{11}B_{1,t-1} + \dots + \alpha_{1p}B_{1,t-p}$  represents the past lag values of the price return;  $\beta_{11}B_{2,t-1} + \dots + \beta_{1p}B_{2,t-p}$  represents the past lag values of the supply return; and  $\varepsilon_{1t}$  represents the disturbance term. The same structure and variables are applied on the right side of equation (2), with a change in the coefficients.

### 2.4. Construction of ARMA-GARCH Model

ARMA-GARCH model combined the Autoregressive moving average (ARMA) model with the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model. In this paper, the ARMA-GARCH model is used to evaluate and forecast the price and supply of bitcoins.

#### 2.4.1 Autoregressive Moving Average (ARMA) Model

ARMA (p, q) model includes AR(p) and MR(p) models. The AR(p) model uses past lag values of price or supply return to forecast future values. while the MR(p) model uses error terms to forecast.

$$y_t = \phi_0 + \sum_{i=1}^p \phi_i x_{t-i} + \alpha_t - \sum_{i=1}^p \theta_i \alpha_{t-i} \quad (5)$$

In equation (2.1), AR(p) and MA(p) are represented by  $\phi_0 + \sum_{i=1}^p \phi_i x_{t-i}$  and  $\alpha_t - \sum_{i=1}^p \theta_i \alpha_{t-i}$  respectively.

#### 2.4.2 Generalized Autoregressive Conditional Heteroskedasticity (GARCH) Model

The generalized Autoregressive Conditional Heteroskedasticity (GARCH) model investigates the pattern of volatility of data. GRACH (1,1) model is the simplified version of the Autoregressive Conditional Heteroskedasticity (ARCH) model. Unlike linear models that assumed variance to be Homoscedasticity, the ARCH model considered the volatility of variance. Derived from the ARCH(p) model, the GARCH (1,1) model was able to forecast variances better compared to other models. In the model used in this paper, another variable volatility of price or supply is added.

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \gamma_1 \sigma_{t-1}^2 + \delta_1 \text{Sigma}_t^2 \quad (6)$$

As shown in equation (2.2), the ARCH model is represented by  $\alpha_0 + \alpha_1 \varepsilon_{t-1}^2$ , the GARCH model is  $\gamma_1 \sigma_{t-1}^2$ , and the volatility variable is represented by  $\delta_1 \text{Sigma}_t^2$ .

## 3. Results and Analysis

### 3.1. Order of VAR Model

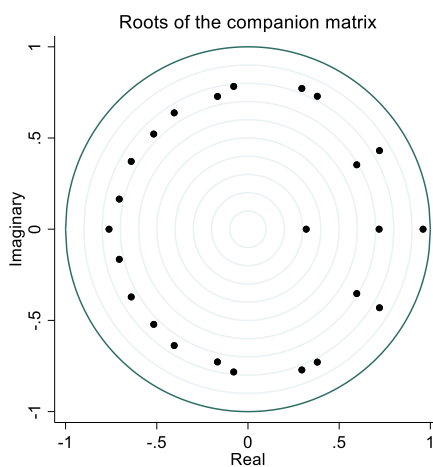
The lag order of the VAR(p) Model used in this paper is found by processing each lag value through the Likelihood-ratio test and other information criteria. The asterisk (\*) will be shown after the value for an appropriate lag value.

**Table 2** VAR(p) model identification

Lag	LL	LR	df	FPE	AIC	HQIC	SBIC
0	32033.7			7.9e-09	-12.9788	-12.9779	-12.9762
1	32507.3	947.290	4	6.5e-09	-13.1691	-13.1663	-13.1612
2	32667.9	321.160	4	6.1e-09	-13.2325	-13.2279	-13.2194
3	32747.6	159.470	4	6.0e-09	-13.2632	-13.2567	-13.2448
4	32814.7	134.190	4	5.8e-09	-13.2888	-13.2805	-13.2651
5	32843.0	56.593	4	5.7e-09	-13.2986	-13.2885	-13.2696
6	32873.0	59.999	4	5.7e-09	-13.3092	-13.2971	-13.2749
7	32911.1	76.140	4	5.6e-09	-13.3230	-13.3091	-13.2834*
8	32921.6	20.977	4	5.6e-09	-13.3256	-13.3099	-13.2808
9	32929.8	16.546	4	5.6e-09	-13.3273	-13.3098	-13.2773
10	32948.0	36.225	4	5.6e-09	-13.3330	-13.3136	-13.2777
11	32966.6	37.204	4	5.5e-09	-13.3390	-13.3177	-13.2783
12	32978.5	23.801*	4	5.5e-09*	-13.3422*	-13.3191*	-13.2763

As shown in Table 2, lag 12 has the greatest number of asterisks. Therefore, for the VAR(p) model used in this paper, the value of p will be 12.

After the specification of the lag value for the VAR model, whether the residuals of the function have autocorrelation and obey normal distribution is tested. It's also important to test whether the VAR model is stationary. Based on Figure 2, because all roots are in the unit circle, VAR (12) is a stable model and there is no need to switch lag value.



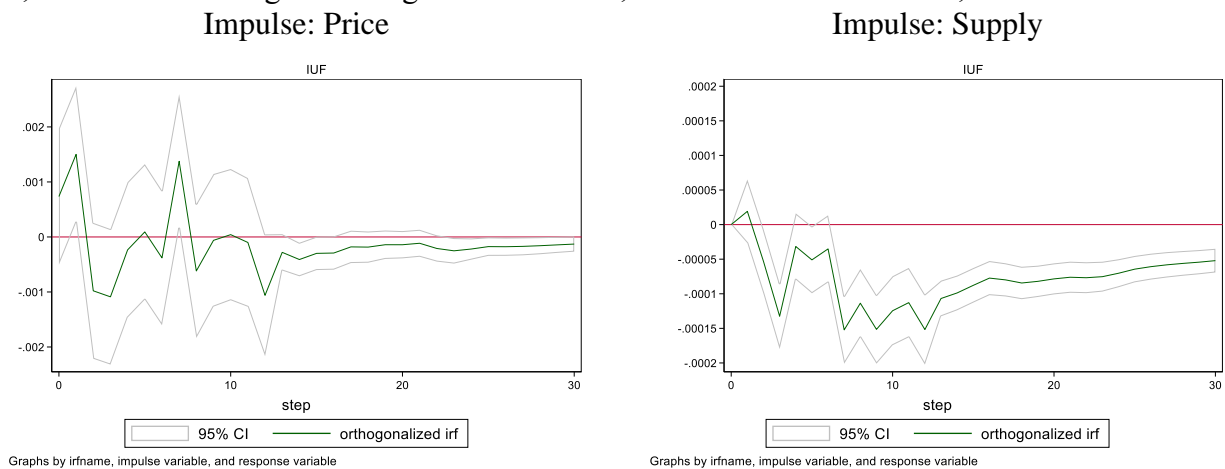
**Figure 2** VAR stability

Photo credit: Original

### 3.2. Impulse Response

The main purpose of this paper is to find correlations between the price of bitcoins(price) and the number of bitcoins held by LTHs(supply). As a result, the Impulse response function (IRF) is used to show the influences of one variable on the other. The impulse response function investigates the amount of change one unit impulse can cause on other variables through time. As shown in Figure 2 impulse and response diagram for the price, when the supply has a unit impulse of 1% at t = 0, the price reaches its greatest positive influence around 0.15% at t = 1. Then the influence switches from positive to negative suddenly between t = 1 to t = 6, hitting its greatest negative influence at t = 3. Immediately after t = 6, it exerts a positive influence around 0.15% at t = 7 dramatically. Nevertheless, the effect is overall negative beyond t = 7, showing that long-term supply doesn't have a significant influence on price.

Also shown in Figure 2 but impulse and response diagram for the supply, when the price has a 1% of unit impulse at  $t = 0$ , the overall effect on supply is negative throughout all time beyond  $t = 2$ .  $t = 6, 8,$  and  $11$  are where greatest negative influences, which is around  $-0.015\%$ , are reached.

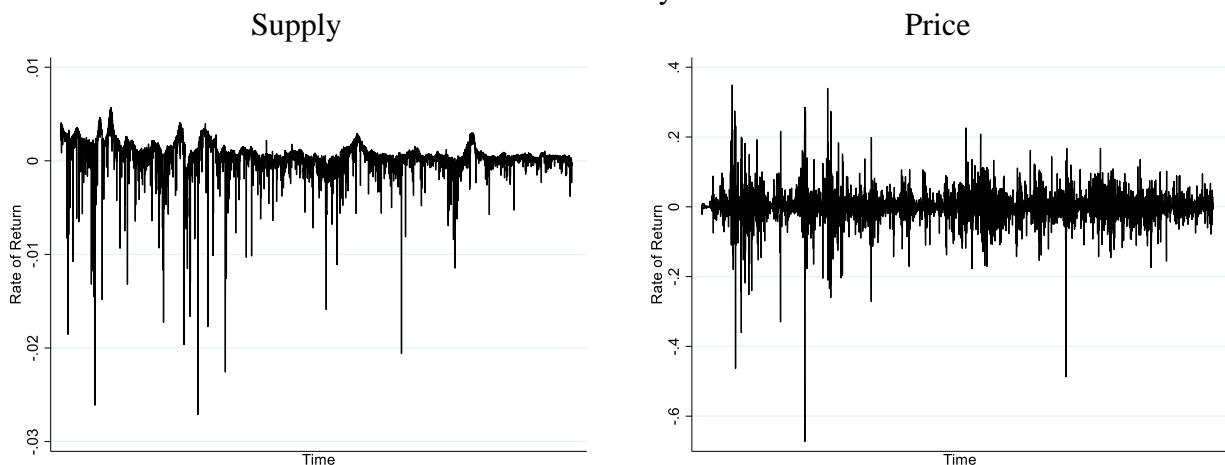


**Figure 3** Impulse and response

Photo credit: Original

### 3.3. Volatility Trend

As said in (2.4.2), the ARMA-GARCH model has a better ability to capture volatility clustering, the behavior of the clustering of observations with large variance or small variance. As shown in Figure 3, the volatility trend of both supply and price seems to be clustered into different parts. This non-constant volatility reveals autoregression conditional heteroskedasticity, thus this paper uses the ARMA-GARCH model to better forecast the volatility of variance.



**Figure 4** Volatility trend

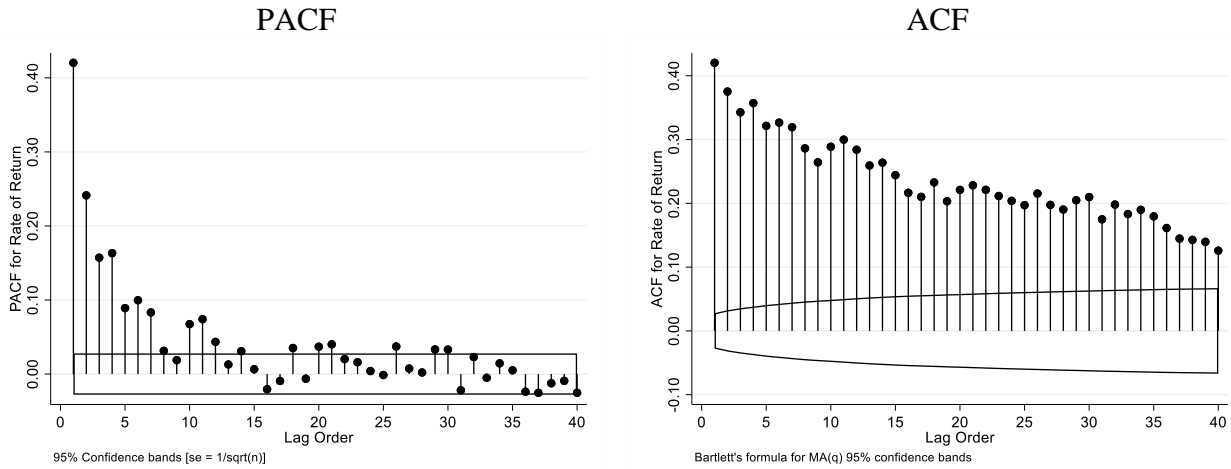
Photo credit: Original

### 3.4. Order of ARMA-GARCH Model

In terms of determining the order of the ARMA-GARCH model, partial autocorrelation function (PACF) and autocorrelation function (ACF) will be used in this paper. In these two functions, the correlations between time series and their different lagged values.

As shown in Figure 4, PACF and ACF are used to determine the order of the ARMA-GARCH model of logarithmic supply return values. In the two graphs, the horizontal black lines indicate the 95%-coefficient intervals. The boundaries are calculated by the formula  $\pm \frac{1.96}{\sqrt{T}}$ , where “T” represents the number of values in the time series. Based on the two diagrams in Figure 4, the first two values

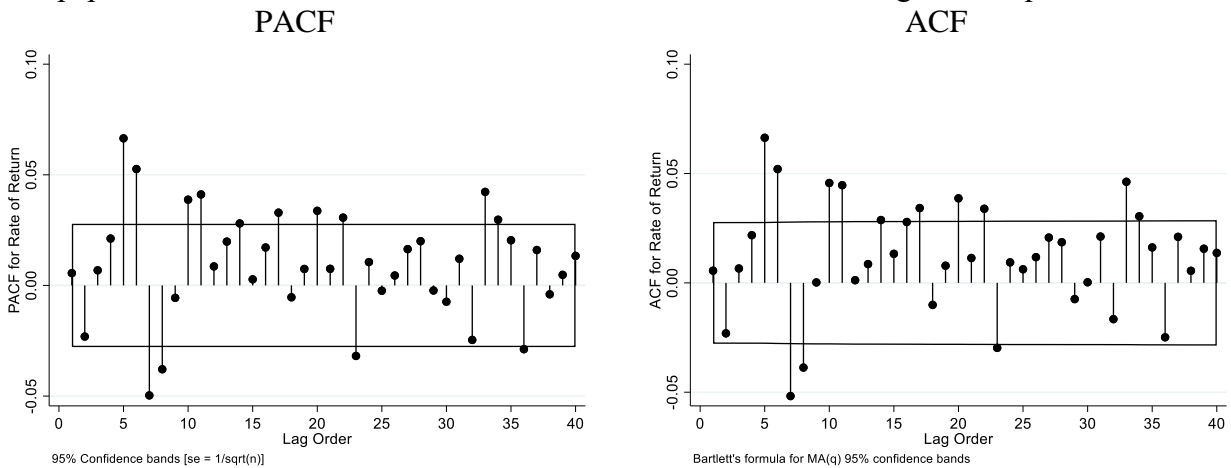
beyond the boundaries of the critical levels are 1. Therefore, p and q values for AR(p) and MA(q) will be order 1.



**Figure 5** PACF and ACF, Supply

Photo credit: Original

Based on Figure 5, AR(p) and MA(q) should both have order 5. However, to simplify the model, in this paper order 1 has been chosen for the ARMA-GARCH model of logarithmic price return values.



**Figure 6** PACF and ACF, Price

Photo credit: Original

### 3.5. Estimated Results of the ARMA-GARCH Model

The ARMA-GARCH estimation results and variance equations for both price and supply are presented in Table 3 below. As shown in row (7) and row(8), all the p-values of coefficients for ARCH and GARCH are significant. This indicates that in both of the ARMA-GARCH models, there is the existence of conditional heteroskedasticity. Shows that there is variance clustering in logarithmic price and supply return values, thus GARCH modeling is appropriate. P-values in both variables Sigma price and Sigma supply are significant. In column 4, the coefficient of Sigma supply is 16379.77, showing a critical influence on the volatility of the price. Similarly, even though the Sigma price has a smaller coefficient, the significance of the p-value still indicates the volatility of price influences the volatility of supply.

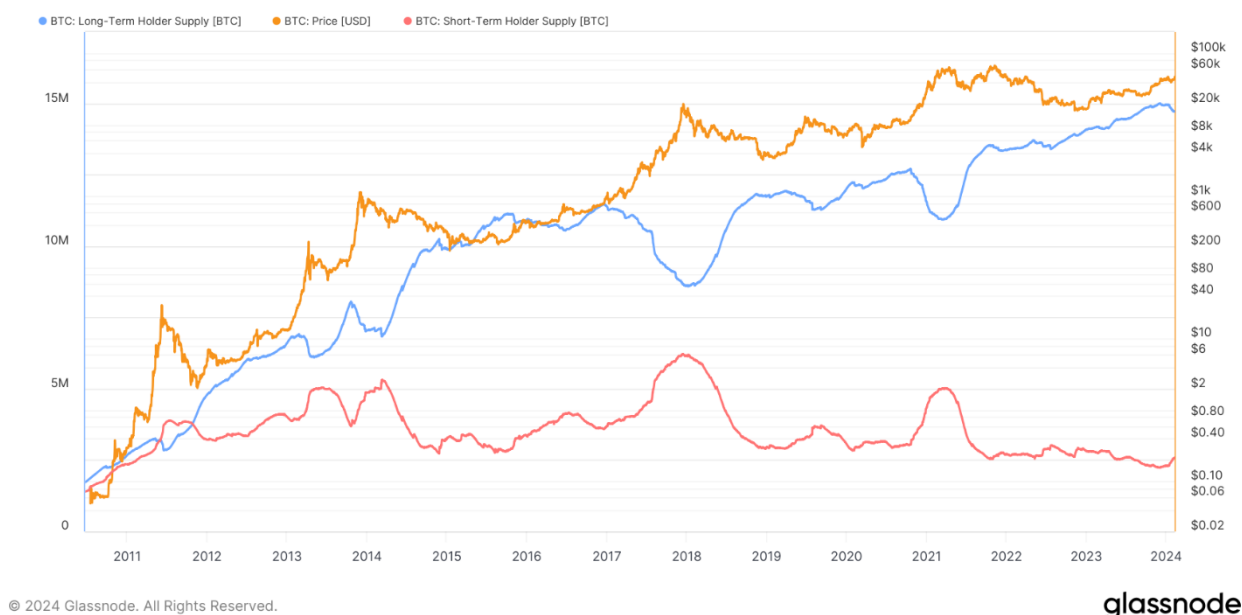
**Table 3** ARMA-GARCH estimation results, variance equation

	(1)		(2)	
	Supply		Price	
	Coef.	P> Z	Coef.	P> Z
Sigma_price	90.89381	0.000***		
Sigma_supply			16379.77	0.000***
ARCH (-1)	0.2045187	0.000***	0.1882387	0.000***
GARCH (-1)	0.4708207	0.000***	0.8017779	0.000***
Constant	-14.13614	0.000***	-9.721062	0.000***

#### 4. Discussion

In this study, the focus of research is on the factors of influence the volatility of Bitcoin price: the number of Bitcoin held by LTHs. Compared to previous studies, this paper continued the usage of the GARCH model for accessing the conditional heteroskedasticity of Bitcoin. In earlier studies, researchers also investigated the application of different GARCH models on Bitcoin analysis. The results of this investigation tend to show EGARCH (1,1) model has better accuracy compared to the GARCH (1,1) and EWMA models in and out of the sample.[12] However, it had been only applied to Bitcoin price but not to the supply of Bitcoin, thus further investigations could be done in the future. This paper can be used by LTHs and STHs in various ways. Managers of financial institutions can use research to evaluate and forecast possible tuning points of Bitcoin price.

BTC: Long-Term Holder Supply, BTC: Price, BTC: Short-Term Holder Supply



**Figure 7** Bitcoin Price (Yellow), Total Supply Held by LTHs (Blue), and Total Supply Held by STHs (Red)

Data source: Glassnode [13]

Photo credit: Glassnode

As mentioned in Section 1, in this market of Bitcoin, the cycle of supply for LTHs had shown a significant relation with the cycle of price change. Nevertheless, as shown in Figure 6, the turning points of STHs seems also highly related to the bull and bear cycle in a way opposite to LTHs [5]. Thus, for STHs, individual speculative investors will have the ability to forecast and learn the behavior of LTHs. Due to the transparency of transactions, quantitative data can be analyzed

compared to the traditional market. Therefore, the result of this paper may also be able to be applied to certain similar traditional markets and used for learning the behavior of LTHs and STHs, like the market of gold.

## 5. Conclusion

The goal of this paper is to investigate the relation between volatility and return of the price of bitcoin(price) and the number of bitcoins held by LTHs(supply). VAR and ARMA-GARCH models are used in this paper to investigate impulse response and evaluate the conditional heteroskedasticity of both variables.

In conclusion, the relation between price and supply is significant. No matter from the perspective of return or volatility, the influence of both on the other is obvious. Based on the 3.2 Impulse response, clear influences from price to supply have been shown. A unit impulse of price creates a long-lasting negative influence on the number of bitcoins held by LTHs. Furthermore, from the ARMA-GARCH model, future values of price and supply were able to be forecasted while considering the existence of conditional heteroskedasticity.

In this paper, the ARMA-GARCH model was applied to assess the influence of one variable's volatility on another. However, based on earlier studies, the GARCH (1,1) model has limitations in investigating the returns and volatility asymmetry of Bitcoin in the long run[14]. Therefore, the EGARCH and TARARCH models were applied. The former, proven by a past study, when combined with the ARMA model can better predict the analysis of the statistical process on Bitcoin returns.

In future studies, STHs should also be considered as another variable that influences Bitcoin price changes, as in Figure 7 the supply of STHs shows a significantly correlated trend with the Bitcoin price. Moreover, it's also possible for researchers to apply the different GARCH-type models to the supply of LTHs and STHs to determine which one suits the best.

## References

- [1] Reserve Bank of Australia. The global financial crisis. (2023-05-26) [2024-02-15]. <https://www.rba.gov.au/education/resources/explainers/the-global-financial-crisis.html>.
- [2] Bitcoin.org. Some bitcoin words you might hear. [2024-02-16]. <https://bitcoin.org/en/vocabulary#confirmation>.
- [3] Yermack D. Is Bitcoin a real currency? An economic appraisal[M]//Handbook of digital currency. Academic Press, 2015: 31-43.
- [4] Dyhrberg A H. Bitcoin, gold and the dollar—A GARCH volatility analysis. Finance Research Letters, 2016, 16: 85-92.
- [5] Schultze-Kraft R. Breaking up on-chain metrics for short- and long-term investors. Glassnode Insights - On-Chain Market Intelligence, Glassnode, (2023-08-18) [2024-02-20]. <https://insights.glassnode.com/sth-lth-sopr-mvrv/>.
- [6] Glassnode. Bitcoin: Total Supply Held by Short-Term Holders [BTC]. Glassnode studio - on-chain market intelligence, [2024-02-12]. <https://studio.glassnode.com/metrics?a=BTC&category=Supply&m=supply.SthSum&mScl=lin&pScl=1og>.
- [7] Cornell B. The relationship between volume and price variability in futures markets. The Journal of Futures Markets (pre-1986), 1981, 1(3): 303.
- [8] Choi H M. Market uncertainty and trading volume around earnings announcements. Finance Research Letters, 2019, 30: 14-22.
- [9] Karpoff J M. The relation between price changes and trading volume: A survey. Journal of Financial and Quantitative Analysis, 1987, 22(1): 109-126.
- [10] Lee C M C, Swaminathan B. Price momentum and trading volume. the Journal of Finance, 2000, 55(5): 2017-2069.

- [11] Grammatikos T, Saunders A. Futures price variability: A test of maturity and volume effects. *Journal of Business*, 1986: 319-330.
- [12] Naimy V Y, Hayek M R. Modelling and predicting the Bitcoin volatility using GARCH models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 2018, 8(3): 197-215.
- [13] Glassnode. Bitcoin: Total Supply Held by Short-Term Holders [BTC]. Glassnode studio - on-chain market intelligence, [2024-02-17].  
<https://studio.glassnode.com/metrics?a=BTC&category=Supply&m=supply.SthSum>
- [14] You Z. Analysis of Cryptocurrency volatility and statistical distributions using ARMA and GARCH-type models. 2019.