

# Statistical Analysis of Realized Volatility of Bitcoin Price using Heterogeneous Autoregressive and Generalized Autoregressive Conditional Heteroskedasticity Models



Aras Jalal Mhamad Karim

Department of Statistic and Informatics - College of Administration & Economics-University of Sulaimani, Sulaimani, Kurdistan Region – Iraq

## ABSTRACT

Bitcoin has recently gained extra attention in the financial industry and the blockchain community in general; it's considered the most popular form of technology. As a result, the purpose of the study is to predict the actual volatility of the bitcoin price using generalized autoregressive conditional heteroskedasticity (GARCH) and heterogeneous autoregressive (HAR). In this research, the researcher attempted to utilize the appropriate statistical methodology, such as GARCH and HAR models. GARCH models were created to address the issue of volatility aggregation, which is the tendency for prices to cluster together as large changes occur. With the GARCH model, we can represent the conditional heteroskedasticity and the fat tail of financial market data. The primary objective was to directly observe and predict the behavior of volatility in time series data. Overall, the model's architecture appears simple and is capable of reproducing the primary characteristics of financial information. The primary concept of this model is that investors with different time frames perceive and respond to different levels of volatility. Sample information about the price of the bitcoin cryptocurrency was distributed worldwide. It includes daily updates of the variable for the time period 31-Jun-17 to 31-Jan-22. The investigation has demonstrated that the HAR model is more effective at predicting variance for this period in comparison to GARCH (1, 1). The result shows that 1 day of previous year's variance estimates and jump estimates have a significant impact on the future variance ( $h = 1$ ).

**Index Terms:** GARCH Model, HAR Model, Forecasting Volatility, Realized Variance, Time Series Analysis

## 1. INTRODUCTION

In recent years, bitcoin has been established as the world's leading cryptocurrency, taking the attention of consumers, businesses, and investors. On the other hand, the topic of forecasting volatility has attracted a lot of attention in the past

decade from many academics and also financial professionals. It has been a subject of great discussion over the years and a lot of research has already been done. One of the most common approaches of modeling volatility indirectly is using ARCH or generalized autoregressive conditional heteroskedasticity (GARCH) models, but nowadays, with the realized measures, it has become possible to directly model volatility. One of the models that directly use the realized measures to forecast volatility is the heterogeneous autoregressive (HAR) model. The major idea of this model is that investors with different time horizons perceive and react to different types of volatility. It is a model that has a simple structure; it is easy to estimate and is able to replicate the main

### Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp136-147

E-ISSN: 2521-4217

P-ISSN: 2521-4209

Copyright © 2025 Karim AJM. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

**Corresponding author's e-mail:** Aras Jalal Mhamad Karim, Department of Statistic and Informatics - College of Administration & Economics-University of Sulaimani, Sulaimani, Kurdistan Region – Iraq. E-mail aras.mhamad@univsul.edu.iq

Received: 14-04-2025

Accepted: 15-06-2025

Published: 18-09-2025

features of financial data, such as in Corsi's study in 2003 [1]. The HAR model is basically an additive cascade of realized volatilities, generated at different time horizons, that follows an autoregressive process. There are a lot of studies in the field of cryptocurrency that has been applied. According to Nathan Reiff's article [2], one of the first attempts to create a cryptocurrency comes from the Netherlands in the late 1980s. At the same time, an American cryptographer, David Chaum, introduced a different electronic cash, *digiCash* [3]. He developed a "blinding formula" to encrypt information to be passed among people. And then some companies applied these fundamentals in the 1990s. In 1998, Wei Dai developed an "anonymous, distributed electronic cash system", called *b-Money* [4]. This system was based on a digital pseudonym used to transfer currency through a decentralized network. The *Bit Gold* proposed by Szabo [5] introduced a proof-of-work system, which is used in some ways in bitcoin's mining network, many academic papers have focused on bitcoin. For Corsi's study in 2003 [1], bitcoin has a position in financial markets and project management between gold and the American dollar with clear advantages to risk averse investors. Zhu *et al.* [6] suggested that some factors, such as Consumer Price Index, Dow Jones Industry Average, and Fed Funds Rate, do have a long-run negative effect on bitcoin price, by applying the vector error correction model. Due to the importance of forecasting, many specialists have developed a variety of time series forecasting models. The framework of historical volatility suggested by Bollerslev [7] on how Realized Variance significantly outperforms GARCH-type models, and the long memory extension presented by Corsi [8] would indicate that this is a highly efficient model. Chung *et al.* [9] applied a HAR model to index options and compared it with IV, which proved significantly higher values of fit. Sea [10] used the HAR-RV model to test its performance on his data. He tested the HAR model against the simple autoregressive (AR) and GARCH (1,1) models. He concluded that the HAR model showed excellent in-sample forecasting performance against another models. According to Vortelinos [11], the HAR model produced the best accurate. The forecast against principal components will combining with neural networks, and GARCH models. From the above studies, the GARCH and HAR models have been widely utilized to enhance the accuracy of the prediction model, especially in the financial field. Therefore, using these models to analyze and forecast of bitcoin cryptocurrency price are suitable. Besides, a modern time series model exists and is applied to analyze bitcoin cryptocurrency, but the specifications and analyzing of bitcoin is always mysterious. For this reason, the present study deals with this problem

by forecasting and analyzing the bitcoin cryptocurrency. Hence, the researcher tries to set up a theoretical model to analyze bitcoin cryptocurrency with the recent record of its data. Hence, the study contribution is building a volatility time series model to estimate and accurate forecast using these models, which are GARCH and HAR. Therefore, the objective of the study is to forecast the realized volatility of the bitcoin cryptocurrency price using GARCH and HAR.

Hence, the next section provides a brief overview of the framework, applying the GARCH and HAR models in the determination of the relationship between variables. In section 3, present the data and derive the time series models utilized in the analysis from the theoretical framework. The conclusions and further discussion of the study results are examined in section 4.

## 2. THEORETICAL FRAMEWORK

### 2.1. Bitcoin

Bitcoin is a peer-to-peer (P2P) electronic cash system introduced in the well-known paper of Nakamoto [12]. The P2P mechanism allows an ownership transfer from one party to another without a third-party intervention (financial institution). Payments can be made over the internet without any control or cost of a central authority for the 1<sup>st</sup> time. Individuals who want to own bitcoins can either run a program on their own computer that implements the bitcoin protocol or create an account on a website that runs bitcoin for its users. The bitcoins are saved in a file called a wallet, which the user may secure and backup. These programs connect to each other over the internet forming P2P networks, making the system resistant to a central attack by M. Crosby et al. [13].

For now, bitcoins are generated through a process of mining. Any member operates as a miner using their computer knowledge to maintain the network. Mining is a computationally process that requires miners to find a solution to a mathematical problem to create a new block into the blockchain. Miners resolve this issue using the proof-of-work concept. This algorithm involves recurrently difficult mathematical problems until getting to a solution. The first miner to find a solution broadcast it to the network to verify it. Once verified, the block is added to the blockchain. Every 10 min on average is found a new answer and a bitcoin is created. The bitcoin protocol is designed to generate a new bitcoin gradually. The difficulty of solving problems is adjusted every 2 weeks at the rate of six blocks per hour.

The size of the reward was initially 50 (genesis block) and it is halved every 4 years, this implies that the number of bitcoins in circulation will never exceed 21 million. Once the last bitcoin is generated, miners will instead be rewarded with transaction fees Lo [13].

## 2.2. GARCH Model

The GARCH model was first introduced by Bollerslev [7]. Back in these days, the concept of realized volatility modeling was not even introduced. At that period, the daily volatility was calculated as the squared daily return without taking into consideration any subintervals. The GARCH model is a conditional volatility model that allows the conditional variance to depend on the previous lags. It is based on the ARCH model by Engle [14], who used it to show that the conditional volatility is affected by volatility clustering. An autoregressive conditionally heteroskedastic (ARCH) model is a time series model with econometric applications that consider the variance of the current error term as a function of the variance of the error conditions of the previous time periods. One of the disadvantages of the ARCH model is that it responds slowly to large, unusual shocks. Thus, the need of an improvement of this model was crucial. Assuming an autoregressive moving average (ARMA) model for the error variance, then the model is a GARCH model. GARCH models were designed to deal with the problem of volatility clustering, which is the phenomenon where large changes in prices tend to cluster together, as A. J. M. Karim and N. M. Ahmed [15]; Botan et al. (2020) [16]; and R. F. Engle [14].

Before describing the GARCH model, the ARCH specification has to be introduced. The following return process has to be specified:

$$r_t = \mu_t + \varepsilon_t \quad \text{with} \quad \varepsilon_t = \varepsilon_t \sqrt{\sigma_t} \quad (1)$$

Where,  $\mu_t$  is a drift term that is explained by the structural model and  $\varepsilon_t$  is an independent shock with zero mean and unit variance, signifying that  $\varepsilon_t$  is normally distributed  $\varepsilon_t \sim Z(0, \sigma)$ . The conditional variance in (1) can be transformed into a time-varying by specifying the ARCH (q) process:

$$\sigma_t = c + \sum_{i=1}^q a_i \varepsilon_{t-i}^2 \quad (2)$$

Where  $c$  is a constant and  $a_i$  is the coefficient for the past squared shocks ( $\varepsilon_t^2$ ). Then the GARCH (p,q) model is derived by adding  $p$  lagged conditional variances, with orders  $p \geq 1$  and  $q \geq 1$ :

$$\sigma_t = c + \sum_{i=1}^q a_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j} \quad (3)$$

Where  $\beta_j$  is the coefficients for the past conditional variances,  $p$  is the past squared error terms, and  $q$  is the past estimated volatility terms. When  $q = 0$ , then the above equation (3) reduces to an autoregressive conditional heteroskedastic (ARCH) model. Given a distribution of  $\varepsilon_t$  in equation (1) and setting  $p = q = 1$ , then the GARCH (1, 1) is derived:

$$\sigma_t = c + a_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1} \quad (4)$$

For which the condition  $c \geq 0$ ,  $a_1 \geq 0$ , and  $\beta_1 \geq 0$  should stand for every positive value of  $\sigma_t$ . Since the GARCH model is non-linear, it cannot be estimated by an OLS regression like the HAR model. Thus, the Gaussian maximum likelihood (GMLE) method should be used for parameter estimation. When assuming normally distributed errors and starting from some parameter vector  $\theta$  and a time series of size  $T$  ( $y_1, y_2, \dots, y_T$ ), the GMLE method calculates the probability density for this specific sample by taking the product over all the marginal conditional probability densities of the observed data. In general, the GARCH model is using the returns to forecast volatility, and it depicts that today's return consists of yesterday's return plus some volatility part and this volatility is what we need. This model is also using a rolling regression method to forecast volatility, by moving one day ahead and leaving 1 day behind for every forecast, which means that the data window size remains stable.

## 2.3. Testing GARCH Effects (Test of Heteroscedasticity)

The availability of ARCH/GARCH effects may give serious model misspecification if they are ignored. Logically, ignoring ARCH effects will give the identification of ARMA models that are over-parameterized. In addition, as in heteroscedasticity, estimation assuming its absence will result in inappropriate standard errors of parameter estimates, which are typically smaller than what they should be. Therefore, it is important to check the presence of GARCH effects in time series modeling according to McLeod and Li [17]; Asraa et al. [18]; and Azhy et al. [19].

Two ways of testing GARCH effects are used. First is to check the Ljung-Box portmanteau Q statistics of  $a_t^2$ .

McLeod and Li show that the sample autocorrelations of  $a_t^2$  have asymptotic variance  $n^{-1}$  and that portmanteau statistics calculated from their distribution is asymptotically Chi-square if  $a_t^2$  are independent. Since the sample

autocorrelations of  $a$  are also pertinent to the identification of a GARCH model for  $a_t^2$ .

The second is the process of checking for conditional heteroscedasticity which is to utilize the Lagrange multiplier test of Engle. Think about the following model of regression for

$$a_t^2 \text{ on } a_{t-j}^2, j=1,2,\dots,m$$

$$a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \dots + \alpha_m a_{t-m}^2 + v_t, t=m+1 \dots n \quad (5)$$

Where  $y_t$  is the error term,  $m$  is a pre-determined positive integer, and  $n$  is the total number of data points in the series. Using the coefficient of determination from (5), Engle demonstrates that, under the null hypothesis  $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_m = 0$ , the variance of  $nR^2$  is approximately distributed according to a chi-square distribution with  $m$  degrees of freedom by Weiss [20]; Nakamoto [12].

## 2.4. Identification of a GARCH Model

If the Ljung-Box statistics and the LaGrange multiplier (LM) test are significant, then conditional heteroscedasticity of  $a_t^2$  is present, and we need to identify an appropriate GARCH model for  $a_t^2$ . However, since the GARCH (1,1) model has been shown to be appropriate in many empirical studies, we may employ the GARCH (1,1) model at the beginning of the analysis. As the model is estimated, diagnostic checking procedures may be followed to see if the GARCH (1,1) model is okay, or if the orders of the GARCH model should be increased or decreased. Instead of using this trial-and-error approach, we may use the following procedure for the definition of a GARCH model for the  $\{a_t^2\}$  series by Lon-Mu [21]; N. M. Ahmed and A. J. M. Karim [22], Dyhrberg [23].

## 2.5. Ljung-Box Q-Statistic

Adding to the visual inspection of the plotted autocorrelation, the Ljung-Box Q-Statistic is used for diagnostic checking by Box and Jenkins [24]. The Ljung-Box Q-Statistic is defined by equation (6)

$$Q^* = n(n+2) \sum_{j=1}^K (n-j)^{-1} r_j^2(a_t) \quad (6)$$

Where  $n$  is the number of observations,  $K$  is the largest degree of freedom used, and  $r_j$  is the sample association function at the  $j$ th degree of freedom of a relevant time series  $a_t$ , for example. The statistical  $r_j$  for is then calculated as:

$$r_j = \frac{\sum_{t=j+1}^n (a_t - \bar{a})(a_{t-j} - \bar{a})}{\sum_{t=1}^n (a_t - \bar{a})^2} \quad (7)$$

The Q-statistic was suggested for testing ARIMA and ARMA models; both the test statistics are determined by the calculation of the sample autocorrelation function for the residuals  $\hat{\varepsilon}_t$  from those models. The similar test statistic based on different calculations using the autocorrelation function will be high benefit for small sample applicability, it is defined as Weiss [20]; M. S. Lo. [25].

$$Q^* = n(n+2) \sum_{j=1}^K \frac{r_j^2(e_t)}{(n-j)} \sim \chi_k^2 \quad (8)$$

where  $k$  is the number of lags considered in the test, and  $k$  is defined by:

$$k = K - m,$$

where  $K$ : Number of lags used in the test.

$m$ : Number of parameters estimated in the mean and variance equations of the GARCH model, and  $r_j^*$  is:

$$r_j^* = \sum_{t=j+1}^n (\hat{a}_t^2 - \bar{a})(\hat{a}_{t-j}^2 - \bar{a}) / \sum_{t=1}^n (\hat{a}_t^2 - \bar{a}) \quad (9)$$

## 2.6. Likelihood Function of GARCH Models

By defining  $\alpha = [\alpha_0, \alpha_1, \dots, \alpha_m, B_1, \dots, B_p, \eta]'$ , the log likelihood functions of  $\alpha$  may be derived under the Normality assumption of  $\varepsilon_t$ . If  $\varepsilon_t$  is assumed to follow a normal distribution. However, practically, there is substantial evidence showing that this assumption may not all the time be satisfactory by Lon-Mu [21].

For the GARCH (1.1) model, the joint density of the observations  $a_1 \dots a_T$  can be calculated as the product of the conditional densities, conditioning on the last observations from M. S. Lo. [25].

$$f_{a_1, \dots, a_T}(a_1, \dots, a_T) = \left\{ \prod_{i=2}^T f_{a_i | a_1, \dots, a_{i-1}}(a_i | a_1, \dots, a_{i-1}) \right\} * f_{a_1}(a_1) \quad (10)$$

Easy to say, the marginal population will decrease as for the ARIMA (1.1) model. For  $k = 2, T$ , the probability of  $a_k$ , given the values  $a_1 \dots a_{(k-1)}$ , is



$$f_{a_k|a_1, \dots, a_{k-1}}(a_k|a_1, \dots, a_{k-1}) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_k^2}} \exp\left\{-\frac{a_k^2}{2\sigma_k^2}\right\} \quad (11)$$

Moreover, the conditional likelihood function given  $a_i$  and  $\sigma_i^2$  is:

$$L(\alpha_0, \alpha_1, B_1) = f_{a_2, \dots, a_T} | a_1 \sigma_1^2(a_2, \dots, a_T | a_1 \sigma_1^2) \\ = \prod_{i=1}^T \frac{1}{\sqrt{2\pi}\sqrt{\sigma_i^{*2}}} \exp\left\{-\frac{a_i^2}{2\sigma_i^{*2}}\right\} \quad (12)$$

Where  $\sigma_i^{*2} = \alpha_0 + \alpha_1 a_{i-1}^2 + B_1$  are obtained recursively. We substitute  $\sigma_i^2$  by its expected value:

$$E(\sigma_i^2) = \frac{\alpha_0}{1 - \alpha_1 - B_1} \quad (13)$$

Using the logarithm and ignoring the constant term, we find that the log likelihood function is:

$$l(\alpha_0, \alpha_1, B_1 | a, \sigma^2) = -\frac{1}{2} \left\{ \sum_{i=2}^T \log \sigma_i^{*2} + \frac{a_i^2}{\sigma_i^{*2}} \right\} \quad (14)$$

Where  $a = (a_1, \dots, a_T)^-$  and  $\sigma^2 = (\sigma_1^2, \dots, \sigma_T^2)$ .

## 2.7. Model Checking of GARCH (r,m)

For a GARCH model, the standardized errors  $\varepsilon_i^\wedge = a_i^\wedge / \sigma_i^\wedge$  are independent and identically distributed random mistakes that are associated with either a standard normal or a non-normal distribution, such as the standardized student-t distribution. As a result, one can assess the effectiveness of a fitted GARCH model by inspecting the series  $\{\varepsilon_i^\wedge\}$ . Specifically, the sample autocorrelations and the Ljung-Box Q statistics of can be utilized to assess the effectiveness of the mean (primordial) equation and the sample autocorrelations and the Ljung-Box Q statistics of can be utilized to assess the validity of the volatility (secondary) equation A.J.M. Karim, and N.M. Ahmed [15]; Lon-Mu [21].

## 2.8. Forecasting the GARCH (1,1) Model

Forecasts of a GARCH model can be found using methods similar to those of an ARMA model. Consider the GARCH (1, 1) model in assume that the forecast origin is n. For a one-step-ahead forecast, we have:

$$\sigma_{n+1}^2 = \alpha_0 + \alpha_1 a_n^2 + B_1 \sigma_n^2 \quad (15)$$

Where  $a_n^2$  and  $\sigma_n^2$  are known at t = n, therefore, the one-step ahead forecast is:

$$\sigma_n^2(1) = \alpha_0 + \alpha_1 a_n^2 + B_1 \sigma_n^2 \quad (16)$$

For multi-step ahead forecasts, we use  $a_i^2 = \sigma_i^2 \varepsilon_i^2$ :

$$\sigma_{i+1}^2 = \alpha_0 + (\alpha_1 + B_1) \sigma_i^2 + \alpha_1 \sigma_i^2 (\varepsilon_i^2 - 1) \quad (17)$$

When t = n+1, the equation becomes:

$$\sigma_{n+2}^2 = \alpha_0 + (\alpha_1 + B_1) \sigma_{n+1}^2 + \alpha_1 \sigma_{n+1}^2 (\varepsilon_{n+1}^2 - 1) \quad (18)$$

Since  $E(\varepsilon_{n+1} - 1 | F_n) = 0$ , the two-step-ahead volatility forecast at the forecast origin n satisfies the equation:

$$\sigma_n^2(2) = \alpha_0 + (\alpha_1 + B_1) \sigma_n^2(1) \quad (19)$$

In general, we have:

$$\sigma_n^2(\ell) = \alpha_0 + (\alpha_1 + B_1) \sigma_n^2(\ell - 1), \quad \ell > 1 \quad (20)$$

This outcome is identical to the result of an ARMA (1, 1) model with an AR polynomial of degree 1 - ( $\alpha_1 + B_1$ ). By repeatedly changing the values in (20), the forward forecast can be written as:

$$\sigma_n^2(\ell) = \frac{\alpha_0 [1 - (\alpha_1 + B_1)^{\ell-1}]}{1 - \alpha_1 - B_1} + (\alpha_1 + B_1)^{\ell-1} \sigma_n^2(1) \quad (21)$$

Therefore:

$$\sigma_n^2(\ell) \rightarrow \frac{\alpha_0}{1 - \alpha_1 - B_1} = \alpha_0^*, \quad \text{as } \ell \rightarrow \infty \quad (22)$$

Provided that  $\alpha_1 + B_1 < 1$

As a result, the multi-step-ahead predictions of volatility made by a GARCH (1,1) model match the unconditional variance of as the horizon for predictions increases to nothing if Var ( $a_t$ ) is present Lon-Mu [21]; Cont [26]; and Nader *et al.* [27].

## 2.9. The HAR-RV Model

The idea of realized variance is based on these assumptions and Andersen *et al.* [28] provide us with an explanation in further detail on how efficient an estimator of volatility the realized volatility is, and moreover, how it outperforms

traditional GARCH-type models. Equation (23) presents the model for estimating a daily Realized Variance (RV), where  $r$  is high-frequency intraday log-returns as described in equation (24).

$$RV_t^{(d)} = \sqrt{\sum_{j=0}^{M-1} r_{t-j}^2} \quad (23)$$

where  $r$  is Log returns (or continuously compounded returns) are approximately equal to normal price returns, but hold significant benefits in simplicity in multi-period returns by Ruppert and Matteson [29]. Log returns are defined:

$$r = \log\left(\frac{P_t}{P_{t-1}}\right), \quad t = 1, \dots, T \quad (24)$$

Based on the Heterogeneous Market Hypothesis (HMH), Corsi [8] proposes the HARRV as a model that will utilize three realized volatility components in an autoregressive manner, which all represent some time-dependent market component for the model. The following equations 25 and 26 consider the RV over the complementing horizons. They are quite simply the average of the daily RV, so for a weekly RV, we simply extend the model as following:

$$RV_t^{(w)} = \frac{1}{5} \left( RV_t^{(d)} + RV_{t-1d}^{(d)} + \dots + RV_{t-4d}^{(d)} \right) \quad (25)$$

Moreover, the same definition for monthly volatility, but over 22 daily periods:

$$RV_t^{(m)} = \frac{1}{22} \left( RV_t^{(d)} + RV_{t-1d}^{(d)} + \dots + RV_{t-21d}^{(d)} \right) \quad (26)$$

The added sum of these three volatilities can be regarded as an additive cascade of volatilities, each representing different components of market volatility. From this, it gets an almost long memory AR type of character (with lags one, five, and 22), but not strictly Corsi [8].

By expanding the expected values and utilizing straightforward recursive substitution, the volatility model will be given by a three-step cascade and has a form of something similar to three AR processes Corsi [8]:

$$\sigma_{t+1}^{(d)} = c + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \omega_{t+1d}^{(d)} \quad (27)$$

Now given (27), all variables are directly observable and available in the data set. The parameters will be able to be estimated through a simple Ordinary Least Squares estimation (OLS) using the Newey-West covariance matrix

estimator. However, due to possible serial correlation, a Newey-West (NW) covariance correction will be applied, since the effects of covariance and autocorrelation must be considered in the estimation.

## 3. DATA ANALYSIS AND RESULTS

### 3.1. Data Description

In this paper, daily observations are used of the bitcoin price, the sample period is June 31, 2017–January 31, 2022, which obtained from the Kaggle website [30], the researcher use R-language to obtain results. Fig. 1 below shows the time series plot of the series during the sample period. Since 2017, the bitcoin price has become more volatile. On October 13<sup>th</sup>, 2017, bitcoin price breaks the \$5,000 for the 1<sup>st</sup> time, on November 28<sup>th</sup>, 2017, the \$10,000, and on December 18<sup>th</sup>, 2017, hits all-time high just below \$20,000. The bitcoin price time series can be observed in Fig. 1, non-linear trend and non-stationarity are the first geometrical properties that are shown in Fig. 1.

To build an appropriate model, the series that are used in analysis must be stationary; therefore, it should check the unit-root structure of the data. Although the above graph gives a rough idea about the stationarity structure of the series, we have applied the Augmented Dickey-Fuller test to the series to test unit roots. Table 1 exhibits the results from the ADF test applied to levels, first differences of the series.

The ADF test results indicate that the variable is non-stationary by not rejecting the null hypothesis of unit-root at the level, but it is stationary after first differencing. Therefore, the researcher uses differenced series in its analysis. Fig. 2 below presents a time series plot of the differenced series.

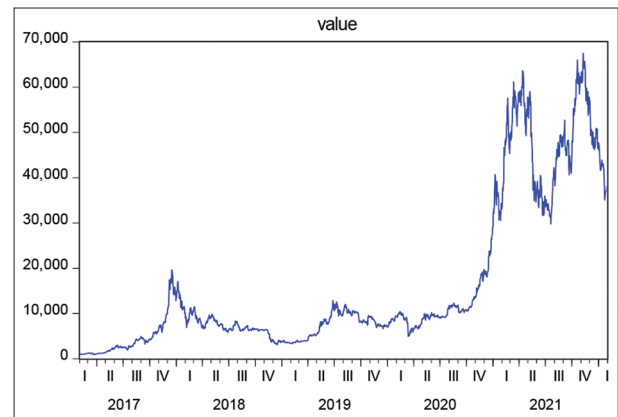


Fig. 1. Time series plots of the variable.

**TABLE 1: Unit root results of the lag variable**

| Variable      | Lags  | Test value | P-value |
|---------------|-------|------------|---------|
| Bitcoin Price | Level | -0.9695    | 0.7659  |
|               | First | -43.9029   | 0.0001  |

After achieving the stationarity condition of the series, we should fit the mean equation model as it is shown below:

From Table 2,  $P$ -value of the constant and lagged variable are less than the statistical significance (0.05), then one can say that the model is significant.

Fig. 3 shows that there is a prolonged period of high volatility from day 1 to the end of 2017, and also, there exists a prolonged period of low volatility from the end of 2018, to the beginning of 2019. This suggests that the residual or error term is conditionally heteroskedastic and it can be represented by the ARCH and GARCH models.

In the series, hypothesis of ARCH has an effect or not on the mean equation. For this purpose, a heteroskedasticity test ARCH have been used to test the following hypothesis below, and its results are shown in Table 3:

- $H_0$ : There is no ARCH effect.
- $H_1$ : There is an ARCH effect.

Table 3 presents  $P$ -value of the heteroskedasticity ARCH test, which is less than the statistical level (0.05), then the hypothesis can be rejected; in another word, there exists an ARCH effect in the series. The researcher achieved two main assumptions of using the GARCH model, which are the stationary of the series and the effect of ARCH in the mean equation model, and then GARCH model can be used to forecast the volatility. The GARCH (1,1) model has been run, the results of its fit are shown in Table 4 below:

From the above table, it is obvious that the estimators of the variance equation are significant, depending on  $P$ -value, which is less than 0.05. The residuals of the GARCH (1,1) model should be tested to find out that the model is suffered from serial correlation of residuals or not for the hypothesis below:

- $H_0$ : There is no serial correlation of residuals.
- $H_1$ : There is serial correlation of residuals.

From Table 5,  $P$ -value of the Q-statistic test for the 36 lags are greater than the statistical level (0.05), then we cannot reject the hypothesis, that is mean there is no serial correlation of residuals.

**TABLE 2: The fit of the mean equation model**

| Variable      | Coefficient | Standard error | t-statistic | P      |
|---------------|-------------|----------------|-------------|--------|
| C             | 5.59E-07    | 3.09E-07       | 1.811371    | 0.0402 |
| Original (-1) | 0.993359    | 0.001206       | 823.5537    | 0.000  |

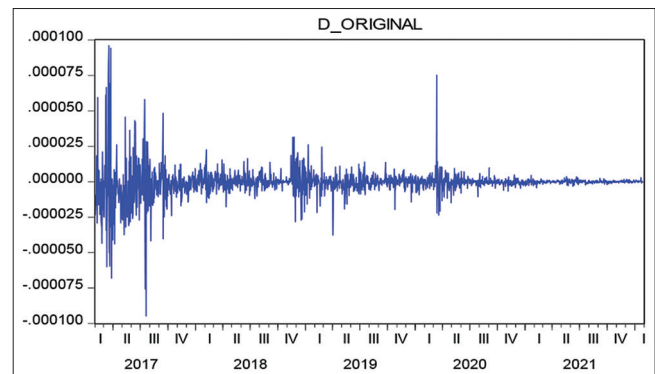
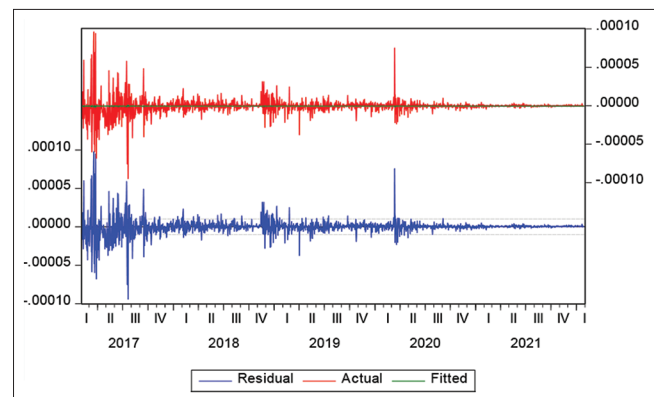
**TABLE 3: The heteroskedasticity test ARCH**

| Test value | P-value |
|------------|---------|
| 214.9139   | 0.000   |

**TABLE 4: The fit of the GARCH (1,1) model**

| Variable          | Coefficient | Standard error | z-statistic | Prob.  |
|-------------------|-------------|----------------|-------------|--------|
| C                 | 3.88E-07    | 1.86E-07       | 2.082135    | 0.0373 |
| Original (-1)     | 0.99336     | 0.000632       | 1570.896    | 0.000  |
| Variance equation |             |                |             |        |
| C                 | 6.35E-12    | 2.59E-13       | 24.52996    | 0.000  |
| RESID (-1)^2      | 0.15        | 0.007405       | 20.2578     | 0.000  |
| GARCH (-1)        | 0.6         | 0.012556       | 47.78632    | 0.000  |

R-squared: 0.9973

**Fig. 2.** Time series plot of the differenced variable.**Fig. 3.** The residuals of the mean equation.

Another test is the heteroskedasticity test to figure out that the postulated model is adequate or not for the following

**TABLE 5: Testing of ACF and PACF**

| Autocorrelation | Partial Correlation | AC | PAC   | Q-Stat | Prob*  |       |
|-----------------|---------------------|----|-------|--------|--------|-------|
|                 |                     | 1  | 0.006 | 0.006  | 0.0659 | 0.797 |
|                 |                     | 2  | 0.006 | 0.006  | 0.1309 | 0.937 |
|                 |                     | 3  | 0.002 | 0.002  | 0.1417 | 0.986 |
|                 |                     | 4  | 0.032 | 0.032  | 2.0313 | 0.730 |
|                 |                     | 5  | 0.028 | 0.028  | 3.4952 | 0.624 |
|                 |                     | 6  | 0.010 | 0.009  | 3.6653 | 0.722 |
|                 |                     | 7  | 0.030 | 0.030  | 5.3633 | 0.616 |
|                 |                     | 8  | 0.010 | 0.009  | 5.5533 | 0.697 |
|                 |                     | 9  | 0.011 | 0.008  | 5.7559 | 0.764 |
|                 |                     | 10 | 0.023 | 0.021  | 6.7321 | 0.750 |
|                 |                     | 11 | 0.031 | 0.029  | 8.5472 | 0.664 |
|                 |                     | 12 | 0.013 | 0.010  | 8.8353 | 0.717 |
|                 |                     | 13 | 0.008 | 0.006  | 8.9489 | 0.777 |
|                 |                     | 14 | 0.018 | 0.015  | 9.5403 | 0.795 |
|                 |                     | 15 | 0.030 | 0.026  | 11.215 | 0.737 |
|                 |                     | 16 | 0.022 | 0.019  | 12.139 | 0.734 |
|                 |                     | 17 | 0.015 | 0.011  | 12.528 | 0.767 |
|                 |                     | 18 | 0.011 | 0.008  | 12.770 | 0.805 |
|                 |                     | 19 | 0.009 | 0.005  | 12.920 | 0.843 |
|                 |                     | 20 | 0.009 | 0.005  | 13.085 | 0.874 |
|                 |                     | 21 | 0.018 | 0.013  | 13.694 | 0.882 |
|                 |                     | 22 | 0.007 | 0.002  | 13.788 | 0.909 |
|                 |                     | 23 | 0.012 | 0.008  | 14.050 | 0.925 |
|                 |                     | 24 | 0.014 | 0.010  | 14.414 | 0.937 |
|                 |                     | 25 | 0.020 | 0.015  | 15.168 | 0.937 |
|                 |                     | 26 | 0.022 | 0.017  | 16.076 | 0.934 |
|                 |                     | 27 | 0.014 | 0.010  | 16.435 | 0.944 |
|                 |                     | 28 | 0.028 | 0.023  | 17.839 | 0.930 |
|                 |                     | 29 | 0.034 | 0.030  | 19.967 | 0.894 |
|                 |                     | 30 | 0.006 | 0.001  | 20.040 | 0.915 |
|                 |                     | 31 | 0.017 | 0.012  | 20.597 | 0.922 |
|                 |                     | 32 | 0.003 | -0.003 | 20.613 | 0.940 |
|                 |                     | 33 | 0.011 | 0.004  | 20.823 | 0.951 |
|                 |                     | 34 | 0.011 | 0.005  | 21.036 | 0.960 |
|                 |                     | 35 | 0.035 | 0.029  | 23.286 | 0.935 |
|                 |                     | 36 | 0.018 | 0.012  | 23.924 | 0.938 |

**TABLE 6: The test of heteroskedasticity for the ARCH effect**

| Test-value | P-value |
|------------|---------|
| 0.065713   | 0.7977  |

hypothesis below, and the results are shown in Table 6 below:

- $H_0$ : ARCH has no effect
- $H_1$ : ARCH has effect.

Table 6 shows the test for heteroskedasticity and it is clear that  $P$ -value of the test is greater than the statistical level (0.05) then cannot be reject the hypothesis, which means that there is no heteroskedasticity.

The final test is a normality test to figure out that the residual of the postulated model is normally distributed or not form the following hypothesis below, and the results are shown in Table 7 below:



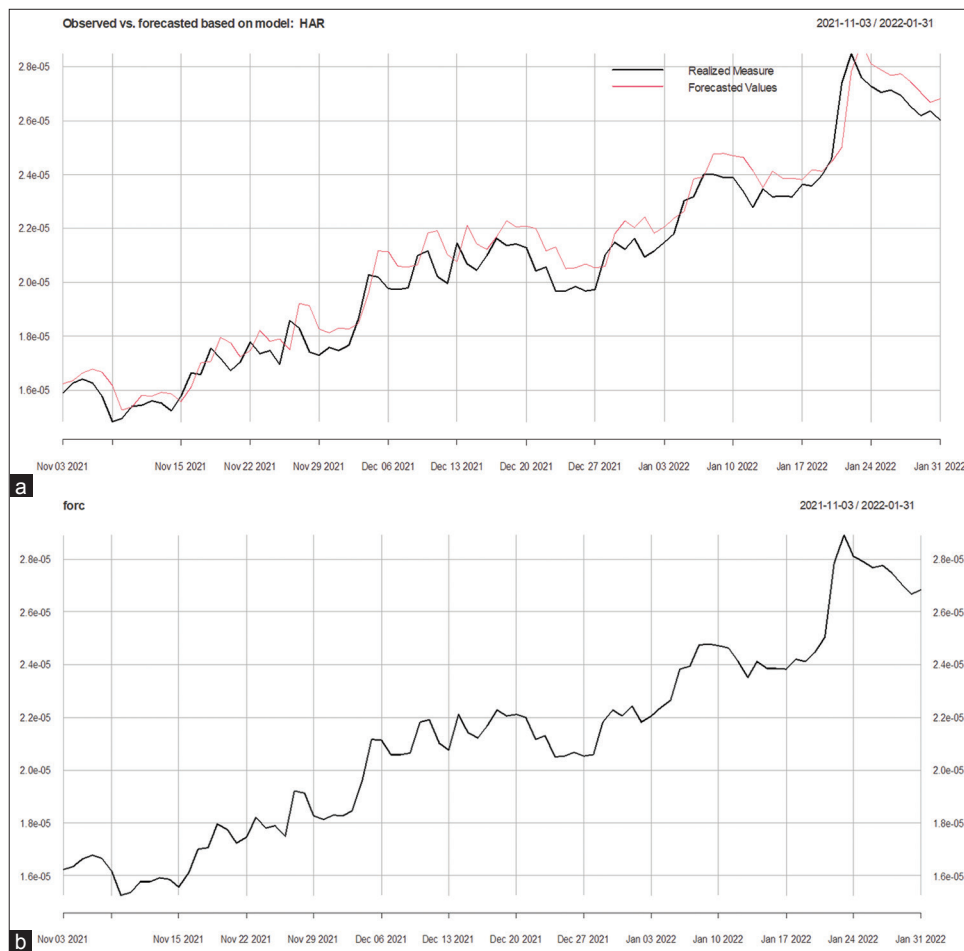


Fig. 4. (a) Observed and forecasted realized volatility (RV), (b) forecasted RV only.

TABLE 7: A normality test for the residual of the model

| Mean  | Median | Max   | Min   | SD   | Skew-ness | Kurt-osis | Jarque Bera | P-value |
|-------|--------|-------|-------|------|-----------|-----------|-------------|---------|
| 0.030 | -0.017 | 16.36 | -8.77 | 1.13 | 1.80      | 32.33     | 66410.98    | 0.000   |

TABLE 8: The forecasted values

| Parameters | Estimate   | Standard error | t value | Pr(> t ) |
|------------|------------|----------------|---------|----------|
| beta0      | 0.00000051 | 0.00000039     | 1.298   | 0.195    |
| beta1      | 0.966      | 0.031          | 31.015  | 0.000    |
| beta2      | 0.042      | 0.039          | 1.083   | 0.279    |
| beta3      | 0.014      | 0.019          | 0.707   | 0.479    |

Multiple R-squared: 0.9968  
F-statistic: 186600 with P-value (0.000)

- $H_0$ : The residual is normal.
- $H_1$ : The residual is not normal.

Table 7 shows the test for residual normality and it is clear that P-value of the test is less than the statistical level (0.05) then one can be reject the hypothesis, which means that the

residual are not normal and this is a good result because residual not normal, which indicate to good model. Moving now to the HAR-RV model, first, the heteroskedasticity test has to be done for the residuals. Using heteroskedasticity test, as shown in Table 3, that there is indeed heteroskedasticity was found. After conducting this test for the HAR-RV model, the realized volatility of yesterday (RV1) was significant at a level of 5%, the realized volatility of past week (RV5) was no again significant, and the realized volatility of past month (RV22) was no again significant at a level of 5%. All the coefficients were positive and the F-test for the model is very large, which is significant. The coefficient for RV1 is more than RV5 and RV22. Table 8 below provides an overview of the coefficients of the estimation.

**TABLE 9: Comparison of postulated models**

| Model | AIC      | BIC    | MSE        | Residual standard error | Log likelihood |
|-------|----------|--------|------------|-------------------------|----------------|
| GARCH | -21.1672 | -21.15 | 1.011E-10  | 1.00E-05                | 19320.1        |
| HAR*  | -36475.5 | -36448 | 6.1419E-13 | 9.829E-06               | 18242.7        |

**TABLE 10: he forecasted values**

| Date              | Forecast | Date              | Forecast | Date             | Forecast |
|-------------------|----------|-------------------|----------|------------------|----------|
| November 03, 2021 | 61619.76 | December 01, 2021 | 54641.91 | January 01, 2022 | 44580.88 |
| November 04, 2021 | 61222.82 | December 02, 2021 | 54747.75 | January 02, 2022 | 45812.86 |
| November 05, 2021 | 60099.69 | December 03, 2021 | 54107.37 | January 03, 2022 | 45356.18 |
| November 06, 2021 | 59618.95 | December 04, 2021 | 50953.54 | January 04, 2022 | 44687.40 |
| November 07, 2021 | 59994.44 | December 05, 2021 | 47245.07 | January 05, 2022 | 44160.97 |
| November 08, 2021 | 61809.40 | December 06, 2021 | 47332.24 | January 06, 2022 | 41972.35 |
| November 09, 2021 | 65592.52 | December 07, 2021 | 48573.66 | January 07, 2022 | 41780.99 |
| November 10, 2021 | 65107.30 | December 08, 2021 | 48597.10 | January 08, 2022 | 40388.64 |
| November 11, 2021 | 63344.30 | December 09, 2021 | 48412.76 | January 09, 2022 | 40347.31 |
| November 12, 2021 | 63388.27 | December 10, 2021 | 45801.95 | January 10, 2022 | 40470.99 |
| November 13, 2021 | 62861.41 | December 11, 2021 | 45644.80 | January 11, 2022 | 40571.47 |
| November 14, 2021 | 63062.32 | December 12, 2021 | 47544.54 | January 12, 2022 | 41413.94 |
| November 15, 2021 | 64183.95 | December 13, 2021 | 48167.93 | January 13, 2022 | 42524.40 |
| November 16, 2021 | 62065.81 | December 14, 2021 | 45213.72 | January 14, 2022 | 41454.08 |
| November 17, 2021 | 58823.18 | December 15, 2021 | 46687.91 | January 15, 2022 | 41922.97 |
| November 18, 2021 | 58644.59 | December 16, 2021 | 47112.76 | January 16, 2022 | 41910.83 |
| November 19, 2021 | 55676.47 | December 17, 2021 | 46145.70 | January 17, 2022 | 41985.74 |
| November 20, 2021 | 56342.73 | December 18, 2021 | 44862.72 | January 18, 2022 | 41329.44 |
| November 21, 2021 | 58032.85 | December 19, 2021 | 45324.90 | January 19, 2022 | 41442.00 |
| November 22, 2021 | 57217.83 | December 20, 2021 | 45247.74 | January 20, 2022 | 40879.03 |
| November 23, 2021 | 54909.94 | December 21, 2021 | 45455.27 | January 21, 2022 | 39958.20 |
| November 24, 2021 | 56151.42 | December 22, 2021 | 47271.15 | January 22, 2022 | 35918.61 |
| November 25, 2021 | 55860.37 | December 23, 2021 | 46954.62 | January 23, 2022 | 34599.21 |
| November 26, 2021 | 57195.02 | December 24, 2021 | 48763.34 | January 24, 2022 | 35582.92 |
| November 27, 2021 | 52061.67 | December 25, 2021 | 48714.33 | January 25, 2022 | 35847.75 |
| November 28, 2021 | 52269.52 | December 26, 2021 | 48353.02 | January 26, 2022 | 36130.89 |
| November 29, 2021 | 54747.87 | December 27, 2021 | 48699.05 | January 27, 2022 | 36036.67 |
| November 30, 2021 | 55152.15 | December 28, 2021 | 48565.47 | January 28, 2022 | 36406.57 |
|                   |          | December 29, 2021 | 45845.78 | January 29, 2022 | 36970.56 |
|                   |          | December 30, 2021 | 44856.14 | January 30, 2022 | 37487.20 |
|                   |          | December 31, 2021 | 45372.75 | January 31, 2022 | 37268.50 |

Table 8 reports the results of the HAR model, where the in-sample forecasting results show consistently that the lagged RV has a strong and persistent positive relationship with future realized variance  $\log(RV_t: t+h)$ , especially in the first forecasting horizons  $h = 1$ . To detect the best model between GARCH and HAR models, one can use the criteria below such shown in Table 9.

Table 9 represents the comparison between models, according to all criteria; the HAR model is the best model than the GARCH model, which means that can be used HAR model to forecast the price of bitcoin cryptocurrency. In addition, the full-sample forecasting assumes the realized variance time series to be stable, so the researcher implements the rolling window method to allow the parameters to change over time, and then more reasonable

comparisons can be obtained. The adaptive method mimics an investor who updates the forecasting model based on the most recent information. The window size  $T$  of the adaptive HAR models employed here is 90 days, that is, models are estimated using past 90-day samples, such as shown in Table 10. Moreover, the model is re-estimated every day. After the re-estimation of each day, the out-of-sample forecasts are performed in horizons  $h = 1; 5; 22$ , spontaneously. The parameters of the daily aggregated realized variances are evolving systematically, which justifies the adaptive forecasting method.

Table 10 reports about the forecast values. These forecasts are obtained by first estimating the parameters of the models on the full sample and then performing a series of static one-step-ahead forecasts. Fig. 4 reports the results for out-

of-sample forecasts of the realized volatility in which the model is re-estimated daily.

Fig. 4 illustrates that the upward and downward risk estimators serve complementary roles in forecasting over time. Specifically, the upward risk coefficients showed a gradual increase from 2017 until early 2021. After that point, they exhibited a stronger upward trend, indicating growing exposure to potential positive risks or gains in the forecasted variable.

## 4. CONCLUSION

As an emerging financial asset, bitcoin has been booming in recent years and has become a major alternative asset for many investors worldwide. This paper studies the realized volatility forecast of the bitcoin cryptocurrency price using GARCH and HAR, based on the data of a 5-year sample period from January 2017 to January 2022. The researcher employs the GARCH and HAR models to study the forecasting properties of bitcoin realized volatility. First of all, a full sample forecasting result reveals that the 1-day lagged realized variance estimators and jump estimators impact the future realized variance significantly across the forecasting horizons  $h = 1$ . Then, the researcher allows the forecasting model to be adaptive with a 90-day rolling window, and who finds that the signed jumps can be a significant predictor of the future realized variance of the longer horizon. The results show that the HAR model seems to successfully achieve the purpose of modeling the behavior of volatility in a very simple and parsimonious way. Moreover, in spite of the simplicity of its structure and estimation, the HAR model shows remarkably good forecasting performance. Based on the out-of-sample forecasting results for the long series of realized volatilities of bitcoin price, the HAR model steadily outperforms the short-memory models at all the time horizons considered (1 day, 1 week, and 2 weeks) and is comparable to the much more complicated and tedious to estimate GARCH model. The study has been found that the HAR model was able to better forecast volatility for this period against GARCH (1, 1).

A good future research suggestion would be that the extensions of these two models could be tested, such as a log HAR model and the 1,2 or 1,3 order of the GARCH model. The EGARCH and TGARCH are two extensions of the GARCH model that can be also tested for their forecasting performance with the HAR model. While high-frequency data enhances the accuracy of realized volatility estimates, it

also introduces microstructure noise. HAR-RV may be more robust to such noise due to its aggregation over multiple time scales, but this can also smooth over meaningful short-term shifts.

## REFERENCES

- [1] F. Corsi. *A Simple long Memory Model of Realized Volatility. Manuscript*. University of Southern Switzerland, Switzerland, 2003.
- [2] N. Reiff. *Were there Cryptocurrencies before Bitcoin*; 2019. Available from: <https://www.investopedia.com/tech/were-there-cryptocurrencies-bitcoin> [Last accessed on 2019 Aug 09].
- [3] D. Chaum. Blind signatures for untraceable payments. In: D. Chaum, R. L. Rivest and A. T. Sherman, editors. *Advances in Cryptology: Proceedings of Crypto 82*. Springer, Germany, pp. 199-203, 1983.
- [4] W. Dai. B-Money; 1998. Available from: <https://www.weidai.com/bmoney.txt> [Last accessed on 2019 Aug 10].
- [5] N. Szabo. Bit Gold; 2008. Available from: <https://unenumerated.blogspot.com/2005/12/bit-gold.html> [Last accessed on 2019 Aug 10].
- [6] Y. Zhu, D. Dickinson and J. Li. Analysis on the influence factors of bitcoin's price based on VEC model. *Financial Innovation*, vol. 3, p. 3, 2017.
- [7] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, vol. 31, no. 3, pp. 307-327, 1986.
- [8] F. Corsi. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, vol. 7, no. 2, pp. 174-196, 2009.
- [9] F. Chung, E. Y. Sun and K. C. Shih. *Do HAR and MIDAS Models Outperform Implied Volatility Model? Evidence from Range-Based Realized Volatility. Manuscript*. Oversea Chinese Institute of Technology, 2008.
- [10] P. Sea. Heterogeneous autoregressive model of the realized volatility: Evidence from Czech stock market. *Advances in Finance and Accounting*, pp. 32-37, 2013.
- [11] D. I. Vortelinos. Forecasting realized volatility: HAR against principal components combining, neural networks, and GARCH. *Research in International Business and Finance*, vol. 39, pp. 824-839, 2017.
- [12] Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," version designed by K. Nordby, Nov. 3, 2020. [Online]. Available: <http://www.KlausNordby.com/bitcoin>
- [13] M. Crosby, P. Nachiappan, P. Pattanayak, S. Verma and V. Kalyanaraman. Blockchain technology: Beyond bitcoin. *Applied Innovation Review*, vol. 7, no. 2, pp. 5-20, 2016.
- [14] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, vol. 55, pp. 987-1007, 1982.
- [15] A. J. M. Karim and N. M. Ahmed. Vector autoregressive integrating moving average (VARIMA) model of COVID-19 pandemic and oil price. *International Journal of Professional Business Review*, vol. 8, no. 1, p. e0988, 2023.
- [16] B. K. Ahmed, S. A. Rahim, B. B. Maarroof and H. A. Taher. Comparison between ARIMA and Fourier ARIMA model to forecast the demand of electricity in Sulaimani Governorate. *Qalaa Zanist*

- Journal*, vol. 5, no. 3, pp. 908-940, 2020.
- [17] A. I. McLeod and W. K. Li. Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis*, vol. 4, pp. 269-273, 1983.
  - [18] A. Asraa, W. T. Shasho, W. Rodeen and H. Tahir. Forecasting the impact of waste on environmental pollution. *International Journal of Sustainable Development and Science*, vol. 1, no. 1, pp. 1-12, 2018.
  - [19] A. A. Aziz, B. M. Shafeeq, R. A. Ahmed and H. A. Taher. Employing recurrent neural networks to forecast the dollar exchange rate in the parallel market of Iraq. *Tikrit Journal of Administrative and Economic Sciences*, vol. 19, no. 62, pp. 531-543, 2023.
  - [20] A. A. Weiss. ARMA models with ARCH errors. *Journal of Time Series Analysis*, vol. 5, pp. 129-143, 1984.
  - [21] L. Lon-Mu. *Time Series Models with Heteroscedasticity*. Vol. 1., Ch. 11. Springer, Germany, 2008.
  - [22] N. M. Ahmed and A. J. M. Karim. Multivariate time series analysis of COVID-19 pandemic and gold price by using error corrections model. *Tikrit Journal of Administration and Economics Sciences*, vol. 19, no. 64, pp. 674-693, 2023.
  - [23] H. Dyhrberg. *Bitcoin, Gold and the Dollar a GARCH Volatility Analysis*. Working Paper Series, no. 15/20, UCD Centre for Economic Research. University College Dublin, Ireland, 2015.
  - [24] E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 1976.
  - [25] M. S. Lo. *Generalized Autoregressive Conditional Heteroscedastic Time Series Models*. Simon Fraser University, Canada, pp. 16-42, 2003.
  - [26] R. Cont. Volatility clustering in financial markets: Empirical facts and agent-based models. In: *Long Memory in Economics*. Springer, Heidelberg, Berlin, pp. 289-309, 2007.
  - [27] R. A. Nader, A. J. Karim and M. M. Hussien. Using artificial neural networks and SPI measure techniques to forecast the risk of drought in Iraq and its impact on environment. *Journal of University of Human Development*, vol. 4, no. 2, pp. 69-77, 2018.
  - [28] T. G. Andersen, T. Bollerslev, F. Diebold and P. Labys. Modelling and forecasting realized volatility. *Econometrica*, vol. 71, no. 2, pp. 579-625, 2003.
  - [29] D. Ruppert and D. S. Matteson. *Statistics and Data Analysis for Financial Engineering*. Springer Texts in Statistics. 1<sup>st</sup> ed. Springer, Germany, 2010.
  - [30] P. Kottarathil. *Bitcoin Historical Dataset*. Kaggle; 2021. Available from: <https://www.kaggle.com/datasets/prasoonkottarathil/btcinud> [Last accessed on 2024 Dec 10].