BITCOIN PRICE FORECASTING UNDER THE INFLUENCES OF NETWORK METRICS AND OTHER FINANCIAL ASSETS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF APPLIED MATHEMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

İSMAİL DAŞ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
FINANCIAL MATHEMATICS

DECEMBER 2022

Approval of the thesis:

BITCOIN PRICE FORECASTING UNDER THE INFLUENCES OF NETWORK METRICS AND OTHER FINANCIAL ASSETS

submitted by **İSMAİL DAŞ** in partial fulfillment of the requirements for the degree of **Master of Science in Financial Mathematics Department, Middle East Technical University** by,

Prof. Dr. A. Sevtap Kestel
Dean, Graduate School of **Applied Mathematics** ________________

Prof. Dr. A. Sevtap Kestel
Head of Department, **Financial Mathematics** ________________

Prof. Dr. A. Sevtap Kestel
Supervisor, **Institute of Applied Mathematics, METU** ________________

Dr. Bilgi Yılmaz
Co-supervisor, **Technical University of Kaiserslautern** ________________

Examining Committee Members:

Prof. Dr. Ömür Uğur
Institute of Applied Mathematics, METU ________________

Prof. Dr. A. Sevtap Kestel
Institute of Applied Mathematics, METU ________________

Assist. Prof. Dr. Furkan Başer
Insurance and Acturial Sciences, Ankara University ________________

Date: ________________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: İSMAİL DAŞ

Signature :

ABSTRACT

BITCOIN PRICE FORECASTING UNDER THE INFLUENCES OF NETWORK METRICS AND OTHER FINANCIAL ASSETS

DAŞ, İSMAİL
M.S., Department of Financial Mathematics
Supervisor : Prof. Dr. A. Sevtap Kestel
Co-Supervisor : Dr. Bilgi Yılmaz

DECEMBER 2022, 56 pages

In 2008, the whitepaper titled "Bitcoin: A Peer-to-Peer Electronic Cash System" introduced an original design of a decentralized digital currency. The innovations, described by the pseudonym Satoshi Nakamoto, have led to the evolution of cryptocurrencies and have attracted significant attention from the media and investors. The price of bitcoin has captured the interest, and the dynamics of bitcoin have become a popular topic in academia. This thesis examines the dynamics of bitcoin prices with several network metrics and financial assets. Three periods chosen by Prophet's change point detection are studied to capture the relationships in different characteristics. Cointegration relationships are analyzed, and according to the results, VAR and VEC models are estimated for granger causality. The cointegration relationship between BTC and hashrate revealed a unidirectional granger-causality from BTC price to hashrate. Additionally, BTC price granger-causes volume in two periods. Finally, ARIMA models are used to forecast Bitcoin and other variables in seven different periods. The results are compared and evaluated by several accuracy measures.

Keywords: ARIMA, Bitcoin, Granger causality test, VAR, Prophet, Cointegration

ÖZ

AĞ METRİKLERİ VE FİNANSAL VARLIKLARIN ETKİSİ ALTINDA BITCOIN FİYAT TAHMİNİ

DAŞ, İSMAİL
Yüksek Lisans, Finansal Matematik Bölümü
Tez Yöneticisi : Prof. Dr. A. Sevtap Kestel
Ortak Tez Yöneticisi : Dr. Bilgi Yılmaz

Aralık 2022, 56 sayfa

2008 yılında, "Bitcoin: A Peer-to-Peer Electronic Cash System" başlıklı makale, merkezi olmayan bir dijital para biriminin özgün bir tasarımını tanıttı. Anonim kimlik Satoshi Nakamoto tarafından duyurulan yenilikler, kripto para birimlerinin evrimine öncülük etti ve medya ve yatırımcılardan büyük ilgi gördü. Bitcoin fiyatı ve bitcoin fiyat dinamikleri de akademide popüler bir konu haline geldi. Bu tez, bitcoin fiyatı ve dinamiklerinin çeşitli network metrikleri ve finansal varlıklar ile bağlantısı incelenmektedir. Prophet'ın değişim noktası tespiti ile tanımlanan iki dönem, farklı özelliklerdeki ilişkileri yakalamak için incelenmiştir. Eşbütünleşme ilişkileri analiz edilmiş ve sonuçlara göre nedensellik analizi için VAR ve VEC modelleri tahmin edilmiştir. Bitcoin fiyatından hacime ve hash oranına tek yönlü granger nedensellik ilişkisi tespit edilmiştir. Son olarak, ARIMA, Bitcoin fiyatını ve diger degiskenleri tahmin etmek için kullanılmış ve sonuçlar, bazı doğruluk ölçümleriyle en doğru tahmin modelini bulmak için karşılaştırılmıştır.

Anahtar Kelimeler: ARIMA, Bitcoin, Granger nedensellik testi, VAR, Prophet, Eşbütünleşme

To my beloved family

ACKNOWLEDGMENTS

I would like to express my deep and sincere gratitude to my thesis supervisor Prof. Dr. A. Sevtap Kestel and co-supervisor Dr. Bilgi Yılmaz for their invaluable guidance and continuous support throughout this thesis. Also, I would like to thank the thesis committe members, Prof. Dr. Ömür Uğur and Dr. Furkan Başer, for their insightful remarks.

I would like to thank my parents, whose love and guidance are with me in whatever I pursue.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

AIC	Akaike Information Criteria
ADF	Augmented Dickey Fuller
ACF	Autocorrelation Function
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average,
BIC	Bayesian Information Criterion
BG	Breusch-Godfrey
DES	Double Exponential Smoothing
DXY	U.S. Dollar Index
ECM	Error Correction Model
ECT	Error Correction Term
ES	Exponential Smoothing
EU ETS	European Union Emissions Trading System
FED	The Federal Reserve
JB	Jarque Bera
MAE	Mean Absolute Error
MA	Moving Average
MAPE	Mean Absolute Percentage Error
MSE	Mean Square Error
PACF	Partial Autocorrelation Function
RMSE	Root Mean Square Error
RSS	Sum of Squared Error
SES	Simple Exponential Smoothing
VAR	Vector Autoregression
VEC	Vector Error Correction

CHAPTER 1

INTRODUCTION

Satoshi Nakamoto published a white paper titled "Bitcoin: A Peer-to-Peer Electronic Cash System" in 2008 [51]. The paper introduced an original design of a digital currency named Bitcoin (BTC), which is a decentralized currency built on a distributed ledger called the blockchain. These innovations, described by the pseudonym Satoshi Nakamoto, have led to the evolution of cryptocurrencies and attracted significant media attention [52]. While the surge in BTC prices has captured the interest of many investors, the dynamics of BTC prices have become a popular topic in academia in recent years.

In the last decade, the attractiveness of BTC as an investment asset has grown significantly. As a consequence, BTC exceptionally climbed in value from less than $1 to over $14,000 between 2010 and 2017 [17]. When its price and overall market valuation soared in the second half of 2017, BTC became one of the most popular topics as a financial investment vehicle. One BTC was worth about $17,000 in the middle of December 2017, and the market capitalization was above $300 billion [28]. The increasing importance of BTC as a financial instrument has resulted in a corresponding increase in the studies on the forecasting of BTC price [2, 3, 16, 31, 38, 43]. While the performance of Prophet and ARIMA is compared in BTC price forecasting [65, 37], machine learning methods are also widely used [50, 56].

BTC secures its blockchain with a process of mining that uses an energy-hungry consensus algorithm called "*proof of work*". The proof of work requires computational power from the network participants to obtain a consensus on data recorded on a blockchain. In this sense, a blockchain can be described as a value-exchange protocol in which miners are rewarded in BTC for their energy expense [48]. The total computational power of the miners is measured by hashrate, which is an input for the cost calculation of miners [33]. In addition, BTC price is critical for all miners to estimate their profits and decide on their mining activities. Therefore, numerous studies have investigated the fundamental relationship between BTC price and hashrate [24, 44, 27, 29].

1.1 Literature Review

Several studies examine the relationship between Bitcoin and various other variables, such as economic indicators, market factors, and news events. Different methodologies are applied to investigate the relationship between BTC and other variables. A number of studies find significant autocorrelation in BTC price and linear dependency with other variables [63, 24, 3, 26, 41, 22, 19]. The concept of Granger causality, introduced by Granger [30], is a common approach to discovering the predictability of the variables. The linear time series models, such as Vector Auto Regression (VAR) and Error Correction models (ECM) are estimated and evaluated for the causality test.

[24] investigated the relationship between BTC price and hashrate in two sub-periods. A cointegration relationship is found between these variables in 11/12/2017 – 4/02/2020, and the unidirectional causality is from BTC price to hashrate. [68] analyzes the influences of some factors on BTC price including oil price, volume, US Dollar index, and stock index by applying a cointegration test with error correction models. The study concludes that all variables have an influence in the long run. They find that gold has the least impact on Bitcoin price, while the dollar index has the greatest impact.

[63] analyzes the influence of the stock index, oil price, and trading volume on BTC price by performing a cointegration analysis through VEC model. Stock price index, oil price, and volume have a long-term relationship with BTC price. The stock price index and oil price have a negative effect on BTC price, while volume has a positive effect. [29] conducts bivariate ECM with BTC price and several variables including hashrate. They find a unidirectional granger-causality from BTC price to hashrate. [42] establishes an empirical model for discovering the dynamics of Bitcoin prices through an error correction model for two periods. S&P 500 index has a positive effect on BTC price, and gold price has a negative effect.

[44] studies between 2013 and 2018, and includes several variables such as hashrate, volume, S&P500, gold, oil, and Google trends to capture the dynamics of BTC price. The study uses Autoregressive Distributed Lag and Generalized Autoregressive Conditional Heteroscedasticity approach and concludes that S&P500 and Google trends affect BTC price. The unidirectional causality, from Bitcoin price to hashrate is captured and evaluated with the mining economics of BTC network. [20] analyzes the causal relationship between Bitcoin attention, measured by Google Trends, and Bitcoin returns by employing the Copula-based Granger Causality in Distribution (CGCD) test. They find a bi-directional causal relationship between Bitcoin attention and Bitcoin returns.

[6] employs a non-parametric causality-in-quantiles test to analyze the causal relationship between trading volume and Bitcoin returns. Volume can predict BTC returns except in bear and bull market regimes while volume cannot predict the volatility of Bitcoin returns. [45] proposes a model to examine the relationship between BTC price and various macroeconomic variables including BTC price, S&P500 volatility index, gold price, and the dollar index. They develop a time-varying cointegration to investigate the equilibrium relationships.

[40] uses wavelet coherence to analyze the co-movement between BTC price and gold. In order to predict the direction of the cryptocurrency market. [25] provide a lower-bound estimate of Bitcoin's intrinsic value by using the cost of the energy used in Bitcoin mining. They include social metrics and studied the bubbles in BTC price. They find positive feedback loops in Bitcoin price and its social activity. [7] finds that the oil price volatility index is an important predictor for both BTC and Gold prices. Ten-year bond yields are also significant for predicting Bitcoin price direction. [18] uses the generalized variance decomposition to understand the sign and severity of spillover shocks from markets such as SP500, gold, and BTC price. The study concludes that cryptocurrencies are unrelated to traditional markets and might provide short-term diversification advantages. [27] finds that S&P 500 index has a negative impact on BTC prices in the long run while showing that positive correlation of twitter activities with BTC prices. In the short-run hashrate has a positive impact on BTC price.

On the other hand, nonlinear models like regime-switching models are also used to capture the nonlinear features among the variables. For example, Markov-switching (MS) AR models, allowing the analysis of causality within the discrete regimes that explain the significant shifts in a time series, are applied in BTC [14]. Markov-Switching GARCH models are used to model the volatility of BTC [15]. Also, Copula methods, investigating the dependency between variables through the joint probabilities of a multivariate distribution, are used for BTC [20, 10].

Given the enormous market capitalization of cryptocurrencies, predicting their prices is a research field with great financial reward potential. Several studies apply efficient forecasting models to predict BTC price. [50] uses Machine Learning (ML) algorithms to predict BTC price. The results show that non-parametric machine learning methods have higher predictive accuracy than regression-based methods. [61] have also used statistical methodologies and ML algorithms to forecast BTC price. [37] applies ARIMA, Prophet, and Multilayer Perceptron (MLP) to forecast the price direction of BTC and MLP provided the best result of accuracy at 54%. [56] explores several algorithms of ML using supervised learning to develop a prediction model and provide an analysis of future market prices.

[31] applies linear regression, logistic regression, SVM, and ANN with the blockchain metrics of Bitcoin, and analyzes the network's influence on Bitcoin price. They predict the price movement with an accuracy of 55%. [3] proposes the prediction model based on ARIMA to evaluate the future value of Bitcoin from 2015 to 2018. They conclude that predicting BTC price is efficient in sub-periods in which the behavior of the time series is stable. In contrast, in the long term, the forecasts of BTC price have large prediction errors. [16] studies with high-dimensional data of BTC, and reveals that logistic regression and linear discriminant analysis achieve an accuracy of 66%. On the other hand, machine learning algorithms give 65.3% accuracy. [38] compares the performances of various deep learning models and their combinations for BTC price forecasting. LSTM-based prediction models are slightly outperformed.

1.2 Aim of the Thesis

Despite the considerable amount of published research about the dynamics of Bitcoin price, there is no consensus on which factors are the drivers of BTC price. While several studies state that BTC price has an impact on hashrate [24, 44, 29], some studies depict different results. Several studies find evidence of Granger causality relationships between Bitcoin and certain variables such as volume, gold, and S&P500 [41, 68, 63, 6, 44]. In this thesis, hashrate and volume are added as the network metrics to extend the previous studies in different time intervals.

BTC price shares common characteristics with commodities such as the mining process and limited supply. Hence, gold, silver, and crude oil are also added to the study. The Dollar index (DXY) and S&P500 are the two important indexes monitored worldwide to interpret a general view of the economy and the financial markets. The allowances of carbon emissions, traded under the European Union Emissions Trading System (EU ETS), encourage reductions in carbon emissions and the use of clean energy. Similarly, there are serious concerns about the fossil energy consumption of BTC mining [4, 60], and BTC is heavily criticized for its carbon emissions. Therefore, the carbon price is also included in the study to investigate the impact of rising carbon prices on BTC price.

BTC price has very distinctive patterns, mostly with high volatility and strong trend, and the changes are quite abrupt due to the regulations, hacking issues, and halving process [9, 8, 54, 47, 49]. So, different relationships may be discovered over different time periods in the studies of Bitcoin price dynamics. Some studies find structural breaks in BTC price and divide the whole period into sub-periods [24, 20], and, interestingly different results are obtained in these sub-periods.

The objective of this thesis is two-fold: The first highlights and evaluates the influence of different variables on BTC price. Through multivariate time series models, this study examines the drivers of BTC price in different time intervals. The second is to make efficient forecasts of BTC price and to compare the performances in different time intervals. The research questions are (i) Does Bitcoin have any causality relationship with the selected variables? (ii) How do these relationships differ in different time horizons? (iii) Which variables have predictive power on BTC prices? (iv) How accurately do $ARIMA$ and ES models predict BTC price in various time periods?

Due to the interpretability limitation of machine learning algorithms, autoregressive models are applied to examine and understand the structure behind the predictive models.

This thesis is expected to contribute to the literature in several points. First, this thesis takes into a wide range of variables, including carbon and oil prices in which the literature is still limited. Also, this thesis examines three distinct periods to compare the relationships in different market conditions. While most studies in the literature include the period until 2020, the upward market movement beginning from the last halving event in 2020 is also included.

The structure of the thesis is as follows. The study's theoretical background and methodology are described in Chapter 2. Empirical results are introduced in Chapter 3. The thesis's key conclusions are summarized in Chapter 4 along with some recommendations for additional research. The final section of the thesis includes the conclusions, along with some recommendations for further research.

CHAPTER 2

METHODOLOGY

This chapter is dedicated to give an intuition to the classical time series modeling methods. We follow [12, 46] closely in this chapter.

A time series is a collection of data points indexed according to time, and its analysis comprises techniques for deriving valuable statistics and other aspects of time series data [46]. Time series models have a variety of forms and can represent various stochastic processes. These models are fitted to time dependent data to investigate the relationships or forecast the future.

The significant contributions [66, 67] lead to establish the foundation for modern time series analysis. These studies primarily focus on the interdependence of time series observations in a simple linear regression model given

$$y_t = \underbrace{x_t^\top \beta}_{\text{explained}} + \underbrace{\varepsilon_t}_{\text{unexplained}}, \qquad t = 1, \ldots, T.$$

The white noise process, frequently used to describe the distribution of the errors in the model, is the basis for most time series models. It is described as a collection of zero-mean, finite-variance random variables that are serially uncorrelated.

2.1 ARIMA

The $ARIMA$ model was primarily introduced in [12]. It is also known as the Box-Jenkins methodology and consists of steps for identifying, estimating, and assessing $ARIMA$ models using time series data. These models are denoted as $ARIMA(p, d, q)$, where p is the order of the autoregression, d is the degree of differencing, and q is the order of the moving average. The AR part of the $ARIMA$ indicates the autoregressive process that is regressed on lagged values, and p is the number of the lags of the autoregressive model. The $AR(p)$ model is written as [12]

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t,$$

where $\varphi_1, \ldots, \varphi_p$ are the finite set of weight parameters, and ε_t is white noise. This model can be written in the equivalent form

$$X_t = \sum_{i=1}^{p} \varphi_i B^i X_t + \varepsilon_t,$$

or

$$\phi[B] X_t = \varepsilon_t,$$

which implies that the autoregressive process can be considered as the result of an all-pole infinite impulse response filter with white noise as its input.

The moving average model of order q is indicated by $MA(q)$, and the model is written as

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} = \mu + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t,$$

where μ is the mean, $\theta_1, \ldots, \theta_q$ are the finite set of weight parameters, and $\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q}$ are white noise. Now, using the backshift operator B, the model can be rewritten in an equivalent form as

$$X_t = \mu + (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t.$$

Conceptually, an MA model is a linear regression of the current value of the series against the past and current error terms. It is assumed that the error terms at each point originate from the same distribution, often a normal distribution, and are independent of one another. They are also considered to have a mean zero and a constant variance.

The $ARMA(p, q)$ model, includes both $AR(p)$ and $MA(q)$ models, and is a linear combination of past values X_t and past errors ε_t as following

$$X_t = \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}.$$

To find the best $ARMA(p, q)$ model, Akaike Information Criterion (AIC) [1], a measure of the goodness-of-fit.

2.2 Exponential Smoothing

Exponential smoothing is a popular and straightforward forecasting technique that works well with discrete time series data. In spite of its simplicity in terms of computational efficiency, it is a powerful forecasting method. Forecasted values are the weighted averages of previous observations. Exponential Smoothing is developed by [13, 34, 64].

Simple Exponential Smoothing (SES) for the time series with no trend and seasonality. In their work, the forecasts depend on the weighted average of previous observations, and are given less weight and have less of an influence on forecasting in the future. In other words, weights decline exponentially as the observations get older [35].

In SES, the forecast y_{t+1} is the weighted average combination of most recent observation y_t with weight α, and the most recent forecast y_t with a weight of $(1-\alpha)$. It is expressed as:

$$s_t = \alpha x_t + (1-\alpha)s_{t-1} = s_{t-1} + \alpha(x_t - s_{t-1}),$$

where α is the smoothing factor, and $0 \leq \alpha \leq 1$. SES is expanded to Double Exponential Smoothing (DES) to account for the trend's effects on the simple exponential smoothing [34]. This approach uses a trend coefficient and a smoothing parameter so that

$$\begin{aligned} \text{Level equation :} \quad & \ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}), \\ \text{Trend equation :} \quad & b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}, \end{aligned}$$

where α $(0 \leq \alpha \leq 1)$ is the data smoothing factor, and β $(0 \leq \beta \leq 1)$ is the trend smoothing factor and, h step ahead forecast equation can be written as,

$$\text{Forecast equation : } \hat{y}_{t+h|t} = \ell_t + hb_t.$$

Holt-Winter Seasonal method is expanded upon by [34, 64] in 1957 in order to account for seasonality. Three smoothing parameters α, β, and γ are needed for level, trend, and seasonal variations, respectively. These equations are as,

$$\begin{aligned} S_t &= \alpha \frac{Y_{t-1}}{I_{t-s}} + (1-\alpha)\left(S_{t-1} + T_{t-1}\right), \\ T_t &= \beta\left(S_t - S_{t-1}\right) + (1-\beta)T_{t-1}, \\ I_t &= \gamma \frac{Y_t}{S_t} + (1-\gamma)I_{t-s}, \end{aligned}$$

where, S is the smoothed observation, T is the trend, I is seasonal index. By using the minimized MSE value, α, β, and γ values are calculated. The k-step ahead forecast is,

$$\hat{Y}_t(k) = \left(S_t + hT_t\right) I_{t+k-s}.$$

2.3 Prophet

The Prophet software, a tool to forecast time series data [59], aims to forecast "at scale," which means a forecasting tool that is automated in nature, making it easier to use when tuning time series methods.

Linear trend with changepoints which is a piece-wise constant rate of growth is provided as,

$$g(t) = \left(k + a(t)^T \delta\right) t + \left(m + a(t)^T \gamma\right),$$

where as k is the growth rate, δ has the rate adjustments, m is the offset parameter, and γ_j is set to $-s_j\delta_j$ to make the function continuous. $a_j(t)$ defined as,

$$a_j(t) = \begin{cases} 1 & \text{if } t \geq s_j \\ 0 & \text{otherwise .} \end{cases}$$

Automatic selection can be done quite naturally with the formulation by putting a sparse prior on δ. A large number of changepoints are specified, and the prior $\delta_j \sim L(0, \tau)$ is used. The parameter τ directly controls the flexibility of the model in altering its rate. Additionally, a sparse prior on the adjustments δ has no impact on the primary growth rate k, so as τ goes to 0 the fit reduces to standard logistic or linear growth.

The trend will have a constant rate when the model's forecast is extrapolated beyond the historical data. By moving the generative model forward, the uncertainty in the predicted trend can be estimated. The generative model for the trend is that there are S changepoints over a history of T points, each of which has a rate change $\delta_j \sim L(0, \tau)$. Future rate changes that emulate those of the past are simulated by replacing τ with a variance inferred from data.

In a fully Bayesian framework this could be done with a hierarchical prior on τ to obtain its posterior, otherwise the maximum likelihood estimate of the rate scale parameter can be used as $\lambda = \frac{1}{S}\sum_{j=1}^{S}|\delta_j|$. Future changepoints are randomly sampled so that their average frequency matches that of the historical changepoints:

$$\forall j > T, \quad \begin{cases} \delta_j = 0 \text{ w.p. } \frac{T-S}{T} \\ \delta_j \sim L(0, \lambda) \text{ w.p. } \frac{S}{T}. \end{cases}$$

Thus, the degree of uncertainty in the predicted trend is estimated by assuming that rate changes will occur with the same average frequency and magnitude as in the past. Once λ has been inferred from the data, this generative model is used to simulate potential future trends. The simulated trends are used to compute uncertainty intervals.

2.4 Cointegration

If two non-stationary time series have the same order of integration and there is a stationary linear combination of these series, then it can be said that these series are cointegrated [23]. Cointegration indicates a common stochastic trend and long-run equilibrium relationship between two variables [23]. In cointegration, it is necessary to use Vector Error Correction Model (VECM). Hence, checking the presence of cointegration is an essential step to determine whether VAR or VECM is appropriate. There are numerous methods for developing cointegration tests, including the Engle-Granger two-step method [23], Johansen test [39], and Phillips-Ouliaris test [53]. This study applies the Engle-Granger two-step method for the cointegration.

If x_t and y_t are non-stationary and order of integration $d = 1$, then a linear combination of them may be stationary for some value of β and u_t. β is estimated first by using ordinary least squares (OLS). Then, Augmented Dickey-Fuller test (ADF) [21] is applied to check unit root of the error term

$$y_t - \beta x_t = u_t,$$

where u_t is the error term. Two $I(1)$ variables variables can be expressed as

$$\begin{aligned} y_{1,t} &= \mu_{y_{1t}} + \upsilon_{y_{1t}} \\ y_{2,t} &= \mu_{y_{2t}} + \upsilon_{y_{2t}}, \end{aligned}$$

where $\mu_{y_{it}}$ represents the trend in variable y_t, and $\upsilon_{i,t}$ is the stationary part. By multiplying y_{1t} by β_1 and y_{2t} by β_2,

$$\begin{aligned} \beta_1 y_{1,t} = \beta_1 \mu_{y_{1t}} + \beta_1 \upsilon_{y_{1t}} \\ \beta_2 y_{2,t} = \beta_2 \mu_{y_{2t}} + \beta_2 \upsilon_{y_{2t}}. \end{aligned}$$

A linear combination of these variables can be expressed as

$$\begin{aligned} \beta_1 y_{1,t} + \beta_2 y_{2,t} &= \beta_1(\mu_{y_{1,t}} + \upsilon_{y_{1,t}}) + \beta_2(\mu_{y_{2,t}} + \upsilon_{y_{2,t}}) \\ &= (\beta_1 \mu_{y_{1,t}} + \beta_2 \mu_{y_{2,t}}) + (\beta_1 \upsilon_{y_{1,t}} + \beta_2 \upsilon_{y_{2,t}}). \end{aligned}$$

If the residuals, $(\beta_1 \upsilon_{y_{1,t}} + \beta_2 \upsilon_{y_{2,t}})$ are stationary, and the linear combination of the variables $\beta_1 y_{1,t} + \beta_2 y_{2,t}$ are also stationary; then their stochastic trend is $-(\beta_1/\beta_2)$ [58].
Therefore, for $y_{1,t}$ and $y_{2,t}$ to be cointegrated,

$$\mu_{y_{1,t}} = \frac{-\beta_2 \mu_{y_{2,t}}}{\beta_1}.$$

If stationarity exists in the residuals, Error Correction Model (ECM) can be estimated such as

$$\begin{aligned} \Delta y_{1,t} &= \gamma_0 + \alpha_1 \left[y_{1,t-1} - \beta_1 y_{2,t-1}\right] + \sum_{i=1}^{K} \zeta_{1,i} \Delta y_{1,t-1} + \sum_{j=1}^{L} \zeta_{2,j} \Delta y_{2,t-1} + \varepsilon_{y_1,t} \\ \Delta y_{2,t} &= \eta_0 + \alpha_2 \left[y_{1,t-1} - \beta_1 y_{2,t-1}\right] + \sum_{i=1}^{K} \xi_{1,i} \Delta y_{2,t-1} + \sum_{j=1}^{L} \xi_{2,j} \Delta y_{1,t-1} + \varepsilon_{y_2,t}, \end{aligned}$$

where β_1 is the cointegrating vector and, α_1 and α_2 are the speed of adjustment parameters.

2.5 Vector Autoregression (VAR)

In time series research, vector autoregressive (VAR) models are frequently used to look at the dynamic relationships between variables [57]. In its basic form, a VAR consists of a set of K endogenous variables $\mathbf{y}_t = (y_{1t}, \ldots, y_{kt}, \ldots, y_{Kt})$ for $k = 1, \ldots K$. The $VAR(p)$ model can be defined as follows after the endogenous variables' p lags are taken into account,

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + CD_t + u_t,$$

where A_i are $(K \times K)$ coefficient matrices for $i = 1, \ldots, p$ and $\mathbf{u}_t$ is a K-dimensional white noise process with time-invariant positive definite covariance matrix, and $\mathbb{E}\left[\mathbf{u}_t \mathbf{u}_t'\right] = \Sigma_{\mathbf{u}}$ such that $\mathbb{E}\left[\mathbf{u}_t\right] = 0$ and, $\Sigma_{\mathbf{u}}$ is a positive definite matrix. A multivariate equation form in $K = 2$ can be written as in the matrix form,

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix},$$

and linear equation system form

$$
\begin{aligned}
y_{1,t} &= c_1 + a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + e_{1,t} \\
y_{2,t} &= c_2 + a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + e_{2,t}.
\end{aligned}
$$

A stable VAR model covariance-stationary when the effect of the shocks, $\mathbf{u}_t$, disappear over time. In this case, the eigenvalues of the companion coefficient matrix λ should be all less than one in absolute value.

$$|\Gamma_1 - \lambda I| < 1$$

To specify the model, the appropriate lag length of the VAR model has to be chosen. AIC or BIC can be used for this purpose as in $ARIMA$ model.

CHAPTER 3

EMPIRICAL ANALYSES

The empirical analyses are done on the variables which are expected to illustrate the outcome of the proposed research questions. The models implemented to the data sets and the forecasts obtained are coded using R and Python software [55, 62]. The variables which are taken into account in frame of forecasting prices are shown in Table 3.1. Hashrate and volume are included as network metrics related to BTC. The influence of commodities, financial market indicators, carbon emission, and environmental index is taken into account to depict their impact on the BTC prices. The mining process of BTC, an energy-consuming process, may cause a relationship with other energy commodities. Carbon, provided by European Union Emissions Trading System (EU ETS), encourages the use of clean energy. Since serious concerns arise about the fossil energy consumption of BTC mining [4, 60], BTC is also criticized for carbon emission, and this similar situation is considered by including carbon emission prices in the study. Additionally, the moves of the Federal Reserve (FED), the central bank of the United States, affect the financial markets worldwide and the Dollar index (DXY). The possible effects of these movements in BTC may also cause a relationship between DXY and BTC.

Table 3.1: Description of the variables

Variable	**Description**	**Source**
Bitcoin price (BTC)	the consolidation of bitcoin prices ($) from major crypto exchanges	Blockchain.com
Hashrate	the estimated number of giga hashes per second performed in Bitcoin network	Blockchain.com
Volume	the total trading volume of Bitcoin	Blockchain.com
Gold	London fixing prices by the London Bullion Market Association (LBMA)	Nasdaq Data Link
Silver	London fixing prices by the London Bullion Market Association (LBMA)	Nasdaq Data Link
Crude Oil	WTI futures price per barrel	Investing.com
Dollar Index (DXY)	the value of the US dollar against major currencies	Investing.com
S&P 500	the stock market index of 500 large companies listed in US	Investing.com
Carbon	the futures price of carbon emissions allowances (EUA) traded on the European Union's Emissions Trading System (ETS)	Investing.com

3.1 Data Description

Most price series increase exponentially, and the multiplicative relationship of the variables is not suitable for linear methods. Hence, it's necessary to make variables additive and linear before modeling with VAR and $ARIMA$. The log transformation scales the data and can help to stabilize variances. The transformation of the natural logarithm is applied to the financial price series. On the other hand, the min-max method is used for the trading volume of BTC.

In many cases, it is necessary to transform time series data before applying any statistical models since the time series data we have usually are not stationary. Further, we should eliminate trends or cycles in the data and use the statistical models for the remaining data. Therefore, we applied the logarithmic transformation for BTC and other price variables. Volume is transformed with the standard min-max normalization defined as

$$y = \frac{x - \min(x)}{\max(x) - \min(x)}.$$

This transformation provides a rescaling of the data so that all values are within the new range of 0 and 1.

The modeling part consists of dividing the data set into training and test subsets, at which we apply the model, estimate the parameters and apply it to the test set to determine if the model fits well using some performance indicators. Based on literature which suggests around 80-20%, we remain within the limits and choose the separation rates as 85-15%. In contrast to BTC, the markets of other financial variables are closed on weekends and holidays. So the time series used in this study are filtered according to the intersection of available observations.

Figure 3.1 shows the visualization of dependent and independent variables, and the increasing trends can be seen in BTC price, hashrate, carbon, and S&P500. Volume, hashrate, and BTC show higher volatility. In last years, although BTC price reached $60,000, the price faced a dramatic fall in 2022. The trading volume has spiked in December 2017 when BTC reached new all-time high levels of around $17,000. The effect of Covid can be seen clearly with sharp movements in crude oil, silver, and S&P500 in Figure 3.1. The price of European carbon emissions and BTC has a significant rise starting from 2021. While DXY moves in a range between 90 and 105 until 2022, the index reached higher levels with the strong trend in this year. S&P500 index has a upward movement with two extreme falls. Silver has an apparent change in level around mid-2020 by rising from $20 to $30. After the logarithmic and min-max transformations, the transformed variables are shown in Figure 3.2. Exponential trends in BTC and carbon have vanished with the logarithmic transformation.

Table 3.2 illustrates the descriptive statistics of the variables. BTC has the highest mean value (8.72), while volume has the lowest one. BTC is also the most volatile variable (1.36), while Dollar Index has the lowest volatile (0.03). If the series is normally distributed, the skewness value should be zero. The kurtosis value, which determines whether a series is heavy-tailed

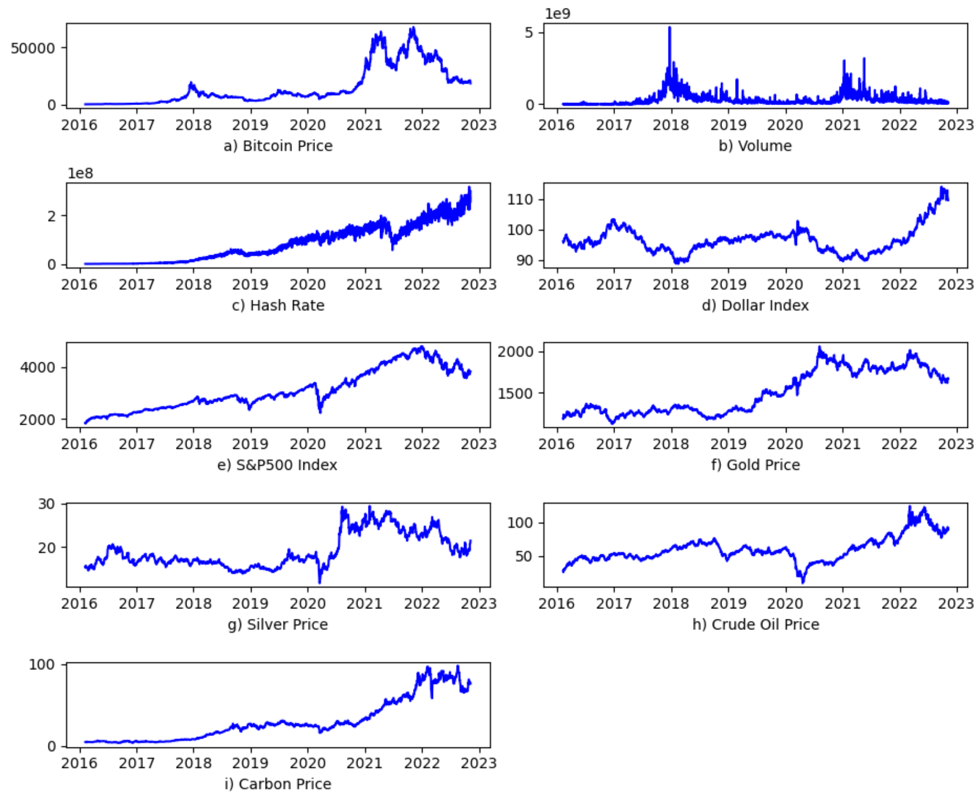


Figure 3.1: Visualization of dependent and independent variables

or light-tailed for a normally distributed series, should have the value of three in normality. The skewness of carbon and S&P500 are close to zero. Volume has the highest kurtosis, which indicates heavier tails than a normal distribution. All variables are very far from the assumption of normality. In Figure 3.3, the differenced series are presented. All time series have mean around zero. While BTC, Dollar index, gold, and carbon have relatively constant variances, volume, crude oil, and S%P500 have high values in some periods.

Table 3.2 shows the descriptive statistics of differenced variables. The mean is zero for all series, and the standard deviations are higher for BTC and its metrics hashrate and volume. Gold, the Dollar index, and S&P 500 are less volatile, with minor standard deviations (0.01). Only gold, volume and DXY have negative skewness, crude oil and and S&P500 are right-skewed with higher positive skewness. Additionally, while all variables have heavy-tails with the positive kurtosis values, Volume, Crude oil, and S&P 500 show extreme values. Jarque Bera statistics are rejected at a 5% significance level for all series, which means they are not normally distributed.

Figure 3.4 graphs the additive seasonal decomposition of the seasonal, trend, and residual of the BTC price. As the figure shows, BTC has a distinct, repeating pattern observed in regular intervals, and the general movement of BTC is increasing. The irregular component consists of the fluctuations in the time series after removing the trend and seasonality components showing periods of low variability in the early and late years of the series. The residuals are serially uncorrelated and have a mean around zero, a white noise.

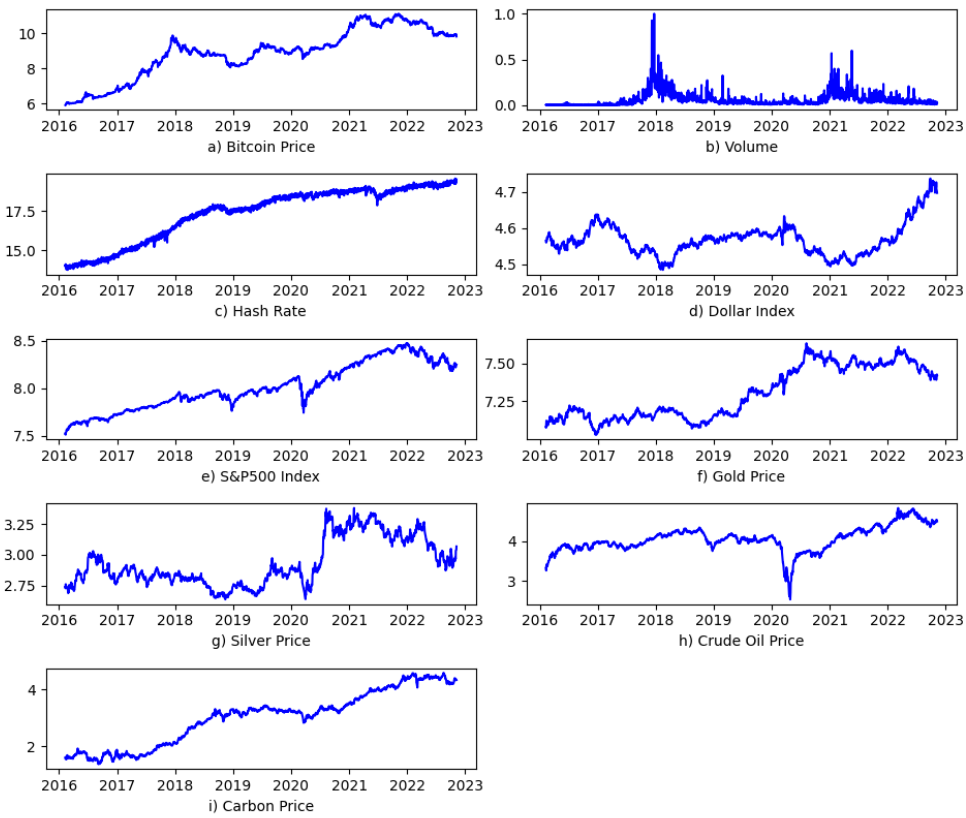


Figure 3.2: Transformed variables

Table 3.2: Descriptive statistics of the variables

	BTC	Hashrate	Volume	Gold	Silver	Crude Oil	Dollar Index	S&P 500	Carbon
Count	1442	1442	1442	1442	1442	1442	1442	1442	1442
Mean	8.89	17.36	0.06	7.27	2.91	3.97	4.55	7.97	2.81
Std. dev.	1.36	1.69	0.08	0.16	0.19	0.25	0.03	0.22	0.83
Min	6.01	13.77	0.00	7.03	2.46	2.30	4.48	7.59	1.38
Max	11.12	19.57	1.00	7.63	3.38	4.43	4.63	8.46	4.49
Skewness	-0.29	0.43	3.73	0.53	0.79	-1.58	-0.01	0.50	-0.20
Kurtosis	-0.57	-1.17	24.87	-1.29	-0.58	5.55	-0.59	-0.53	-1.18
JB test	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

(a) Transformed variables

	BTC	Hashrate	Volume	Gold	Silver	Crude Oil	Dollar Index	S&P 500	Carbon
Count	1441	1441	1441	1441	1441	1441	1441	1441	1441
Mean	-0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Std. dev.	0.05	0.12	0.05	0.01	0.02	0.03	0.01	0.01	0.03
Min	-0.26	-0.45	-0.62	-0.07	-0.09	-0.31	-0.02	-0.09	-0.20
Max	0.27	0.49	0.69	0.05	0.12	0.36	0.02	0.13	0.16
Skewness	0.65	0.16	-0.09	-0.08	0.77	1.96	-0.10	1.14	0.47
Kurtosis	9.01	5.69	45.08	5.08	7.80	34.07	1.56	23.40	4.07
JB test	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

(b) Differenced variables

Table 3.3 presents the correlation matrix to illustrate the linear relation between the independent and dependent variables. The dependent variable has the highest correlation with S&P500 (0.93). Although BTC and S&P500 are two distinct financial assets with different

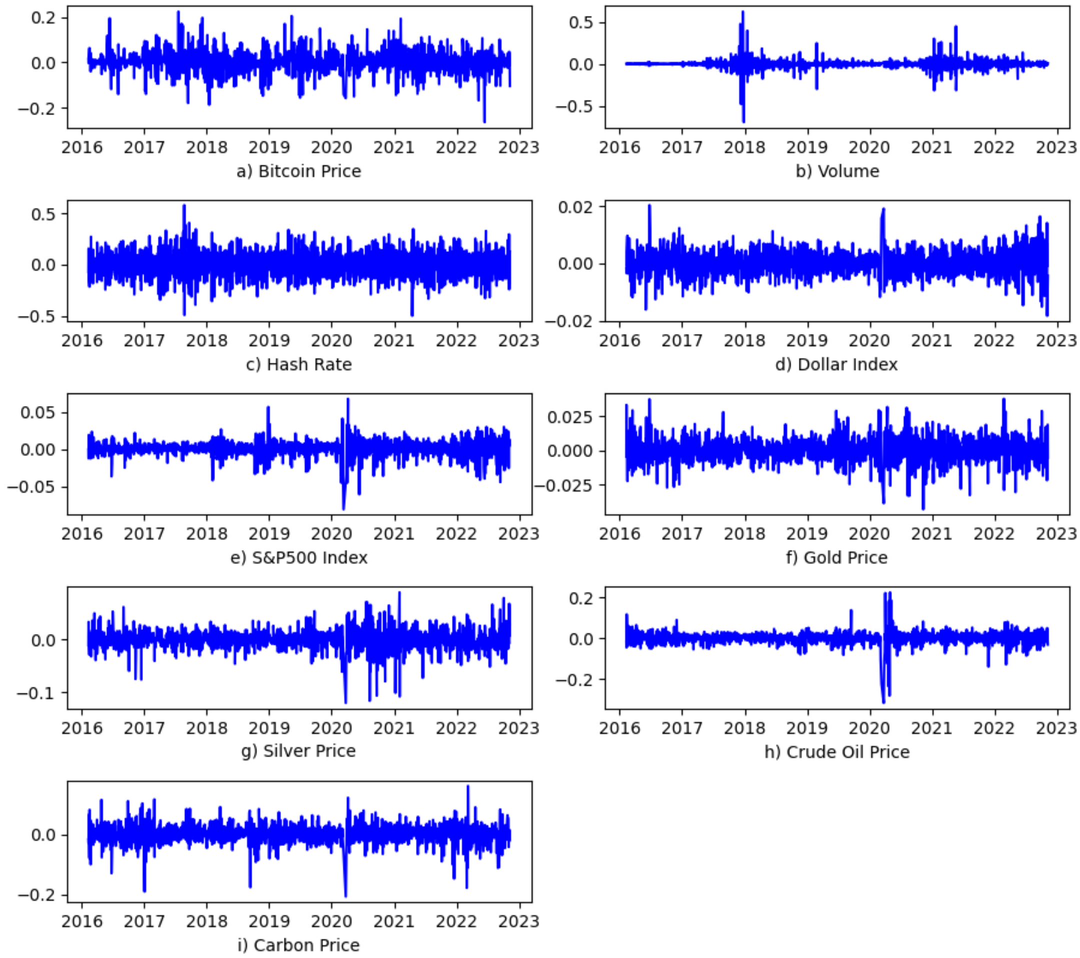


Figure 3.3: Differenced variables

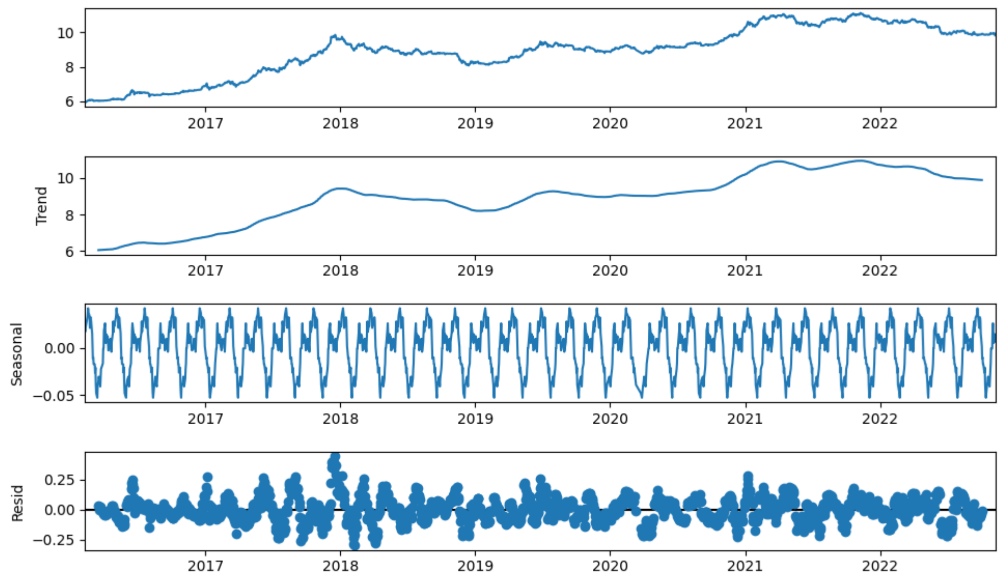


Figure 3.4: Seasonal decomposition of BTC

characteristics, they have similar trends starting from 2020. One reason for this common movement is the positive investor sentiment by the recovery from the pandemic in mid-2020.

Hashrate and BTC also have a strong correlation with 0.90. This strong correlation between BTC and hashrate investigated in many studies [24, 44, 68]. The economics of the mining process in BTC network is frequently discussed in this aspect. Additionally, Bitcoin has significant correlations with the other variables except for the Dollar index (-0.11). DXY is a measure of the value of the dollar relative to several major currencies, and a higher DXY implies a stronger dollar. Hence, the value of the DXY can negatively impact commodities priced in dollars, such as gold, silver, and crude oil. DXY has smaller correlation values than the other variables.

While carbon price is a measure of the cost of carbon dioxide emission, BTC price and hashrate are related to the mining process which leads to carbon emissions. There is also a strong correlation between these variables. Various studies are motivated to investigate these strong relationships of BTC with other financial assets [26, 6, 68]. The correlation between S&P500 and Hashrate (0.86), S&P500 and Gold (0.82), S&P500 and Silver (0.70), S&P500 and Carbon, and Silver and Gold (0.82) are relatively high.

Table 3.3: Correlation coefficients among the variables

	BTC	Hashrate	Volume	Gold	Silver	Crude Oil	Dollar Index	S&P 500
Hashrate	**0.90**	1						
Volume	**0.74**	**0.62**	1					
Gold	**0.73**	**0.73**	0.30	1				
Silver	**0.60**	0.40	0.17	**0.82**	1			
Crude Oil	**0.55**	0.19	0.36	0.26	0.28	1		
Dollar Index	-0.11	0.08	-0.42	-0.37	0.22	0.22	1	
S&P 500	**0.93**	**0.86**	0.19	**0.82**	0.70	0.39	0.02	1
Carbon	**0.87**	**0.94**	**0.54**	**0.73**	0.47	0.34	-0.29	**0.82**

3.1.1 Time Periods

In this study, further analyses examine the relationship between variables over a certain period of time. However, the relationship between these variables may not be constant across the entire period. The underlying relationship between the variables in the model can change due to some external factors. The standard framework for structural breaks, provided by [5], is widely used to detect these structural changes. Also, Prophet [59] offers the detection of change points in the series. In this study, Prophet is used to capture abrupt changes linearly in BTC price.

Table 3.4: Changepoints in BTC price

1	2	3	4	5	6	7
2016-12-20	2018-01-21	2018-12-31	2019-06-25	2020-03-25	2021-02-18	2022-04-14

Prophet takes parameters such as the number of change points, the prior scale, and the range to estimate change points. Here, the number of change points is specified as seven at the beginning, and the changepoints can be seen in BTC price in Figure 3.5. To capture the downside movement this year, the range is set as 0.90. Table 3.4 shows the seven change points in BTC price from 2016 to the end of 2022.

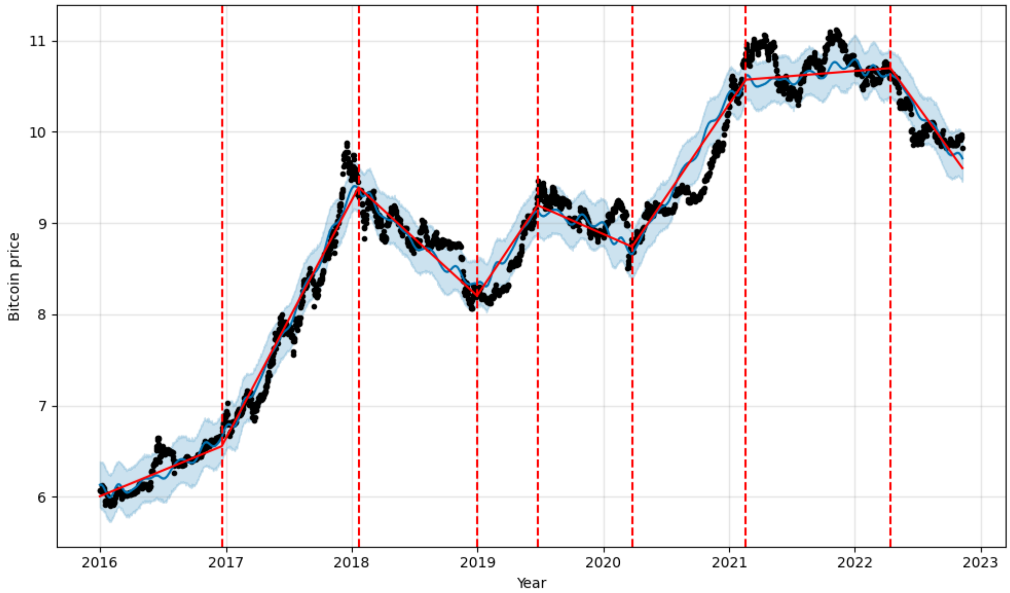


Figure 3.5: Changepoints in BTC Price

The rate of the change points is presented in Figure 3.6. The highest change is at the beginning of 2018 when the long bull run of BTC ended. One critical finding is that the continuous upside trend from the beginning ended dramatically in March 2022. Second, the bull run, beginning in March 2020, is well captured by Prophet.

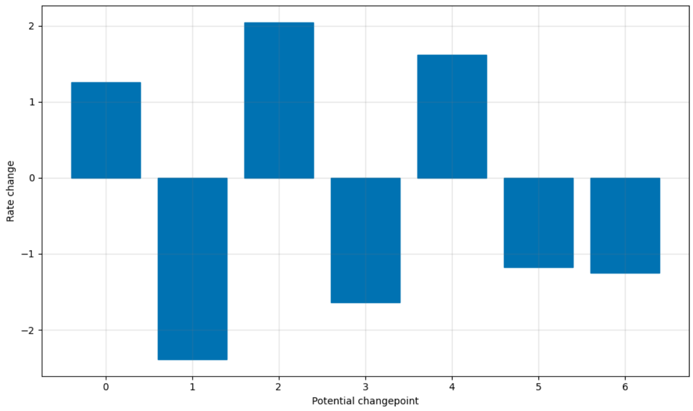


Figure 3.6: Rate of the changepoints

In this study, according to three different market conditions which are flat, up, and down, six sub-periods are specified for further analyses with $ARIMA$ by selecting two periods from each market condition. The interval from 2016 to the end of 2022 which covers all periods of the study is also included as the first period. All intervals are listed in Table 3.5.

Table 3.5: Time periods

Univariate $ARIMA$	Time intervals	Multivariate models	Time intervals
Period 1: All	2016-12-20 - 2022-11-08	Period 2: Up	2020-03-25 - 2021-02-18
Period 2: Up	2020-03-25 - 2021-02-18	Period 4: Down	2022-04-14 - 2022-11-08
Period 3: Up	2016-12-20 - 2018-01-21	Period 6: Flat	2021-02-19 - 2021-09-19
Period 4: Down	2022-04-14 - 2022-11-08		
Period 5: Down	2018-01-21 - 2018-12-31		
Period 6: Flat	2021-02-19 - 2021-09-19		
Period 7: Flat	2021-09-20 - 2022-04-14		

The second period includes the upside movement beginning from March 2020 to February 2021 while the other up period 3 covers the bull-run in 2017. Periods 4 and 5 include the downside movement of BTC price. The fourth includes the last downside movement of BTC beginning from April 2022 in which the price is around $60k. Although the significant deviation, BTC price has no level change between February 2021 and April 2022. So this interval is described as flat and divided into two sub-periods as Periods 6 and 7 in order to make comparable analyses in the same market condition.

According to the results of the stationarity tests and $ARIMA$ models in the following chapter, this study has selected one period from each market condition to investigate the relationship between the variables with a multivariate framework. The time periods of the multivariate modeling can be found in Table 3.5.

3.1.2 Stationarity

The Augmented Dickey-Fuller test [21] is used to verify whether the time series is stationary, and the test results for period 1 are presented in Table 3.6. First, the test equation includes both a constant and time trend. If the null hypothesis of the test cannot be rejected, then the F statistics of the coefficients are checked. According to the statistics, the time trend component may be removed from the test equation.

In case of a deterministic time trend in variables, the test may indicate stationarity with a trend component in ADF equation. The null hypothesis of the ADF test cannot be rejected for the preprocessed series in Period 1, and as seen in Table 3.6, the series has a clear trend, especially BTC, hashrate, carbon, and S&P500.

The dollar index, crude oil, and silver have extreme volatility at some points. However, after the differencing operation, all preprocessed variables have p-values close to zero, indicating stationarity in Table 3.6. The test results of other periods can be found in Tables 3.7 3.8. In Period 2, according to both normality tests, hashrate is normally distributed. In Period 3, while gold and crude oil are normally distributed according to both tests, silver and $S\&P500$ are normal only by JB test. In Period 4, hashrate, gold, and $S\&P500$ are normally distributed. However, silver, crude oil, and DXY are also normally distributed by JB test.

Table 3.6: Stationarity test results of Period 1

	Transformed						Differenced		
	ADF (c)			ADF (ct)			ADF (c)		
Variables	ADF-Stat	p-value	n-lags	ADF-Stat	p-value	n-lags	ADF-Stat	p-value	n-lags
BTCPrice	-0,991	0,757	1	-1,648	0,773	1	-37,091	0,000	1
Hashrate	-2,448	0,129	7	-0,293	0,990	7	-15,755	0,000	10
Volume	-2,520	0,111	10	-2,538	0,309	10	-11,361	0,000	9
Gold	-0,834	0,809	0	-1,929	0,640	0	-36,412	0,000	3
Silver	-1,419	0,573	2	-2,037	0,581	2	-19,393	0,000	1
Crudeoil	-2,833	0,053	8	-2,615	0.273	8	-8,101	0,000	7
DXY	-2,154	0,223	0	-2,310	0,428	0	-37,351	0,000	0
SP500	-0,086	0,951	3	-2,164	0,510	3	-9,709	0,000	2
Carbon	-0,499	0,892	0	-2,210	0,484	0	-39,169	0,000	0

Table 3.7: Stationarity and normality of transformed variables: Periods 2-4

			JB Test		Shapiro Test		ADF ('c')			ADF ('ct')		
	Skew	Kurtosis	stats	p-value	stats	p-value	ADF-Stat	p-value	n-lags	ADF-Stat	p-value	**n-lags**
	Period 2											
BTCPrice	0,759	0,260	18,068	0,000	0,931	0,000	1,340	0,997	1	0,058	0,995	1
Hashrate	-0,325	-0,189	3,605	0,165	0,989	0,141	-2,871	0,049	2	0.389	0,798	2
Volume	2,206	7,044	512,162	0,000	0,794	0,000	-2,091	0,248	4	-2,668	0,250	4
Gold	-0,342	-0,749	8,087	0,018	0,962	0,000	-2,118	0,237	0	-1,854	0,678	0
Silver	-0,510	-1,159	18,471	0,000	0,891	0,000	-1,754	0,404	1	-1,483	0,835	1
Crudeoil	-1,913	3,083	180,916	0,000	0,737	0,000	-1,670	0,446	4	-2,399	0,380	4
DXY	0,340	-1,074	12,589	0,002	0,925	0,000	-0,175	0,941	0	-2,500	0,328	0
SP500	-0,710	0,039	15,480	0,000	0,951	0,000	-2,739	0,068	4	-3,737	0,020	0
Carbon	-0,657	-0,284	13,967	0,001	0,949	0,000	-2,136	0,230	1	-2,402	0,379	1
	Period 3											
BTCPrice	0,135	-1,287	15,725	0,000	0,929887	0,000	0.126	0.967	0	-3,002	0,131	0
Hashrate	0,179	-1,065	11,563	0,003	0,96277	0,000	-1,128	0.703	3	-4,922	0,001	3
Volume	2,300	6,594	566,157	0,000	0,748736	0,000	2,216	0.998	13	0,700	1,000	13
Gold	-0,239	0,257	2,525	0,283	0,991145	0,204	-2,130	0.232	12	-2,922	0,155	12
Silver	-0,014	-0,330	1,126	0,570	0,986736	0,039	-2,787	0.060	3	-3,007	0,130	3
Crudeoil	-0,164	-0,474	3,169	0,205	0,987785	0,058	-1,748	0.406	0	-1,609	0,788	0
DXY	-0,018	-1,482	19,890	0,000	0,916989	0,000	-1,266	0.644	0	-1,846	0,682	0
SP500	0,007	-0,666	4,205	0,122	0,975284	0,001	-1,479	0.543	0	-3,643	0,026	0
Carbon	0,774	-0,787	27,362	0,000	0,869984	0,000	-0.126	0.950	2	-2,291	0,438	2
	Period 4											
BTCPrice	0,981	-0,242	18,825	0,000	0,849	0,000	-1,941	0,313	0	-2,109	0,541	0
Hashrate	0,065	-0,001	0,093	0,955	0,995	0,939	-0,861	0,801	4	-1,105	0,928	4
Volume	2,917	9,726	585,443	0,000	0,677	0,000	-5,548	0,000	0	-4,894	0,000	5
Gold	0,132	-0,642	2,511	0,285	0,981	0,097	-1,794	0,384	0	-2,932	0,152	0
Silver	0,539	-0,119	5,700	0,058	0,963	0,002	-2,789	0,060	0	-2,718	0,229	7
Crudeoil	-0,030	-0,941	4,475	0,107	0,976	0,031	-1,338	0,611	0	-2,502	0,327	0
DXY	0,300	-0,648	3,937	0,140	0,977	0,037	-0,910	0,785	0	-2,733	0,223	0
SP500	0,031	-0,609	1,997	0,368	0,986	0,266	-1,778	0,391	0	-1,965	0,621	0
Carbon	-0,731	-0,321	10,886	0,004	0,916	0,000	-1,203	0,672	0	-2,319	0,424	0

BTC price is trend stationary in Period 5. While $S\&P500$ is trend-stationary in Periods 2 and 3, hashrate is trend-stationary in Periods 3 and 7. On the other hand, volume is stationary in Periods 4, 6 and 7. Silver is only stationary in Period 7.

3.2 ARIMA Models

The $ARIMA$ model is established with the Box-Jenkins procedure. In the first part of, transformations and differencing are involved in data preparation [12]. In a series where the variance varies with the level, data transformations (such as logarithms) can help stabilize the variance [11]. Then the data are differenced until there are no apparent patterns such as trend

Table 3.8: Stationarity and normality of transformed variables: Periods 5-7

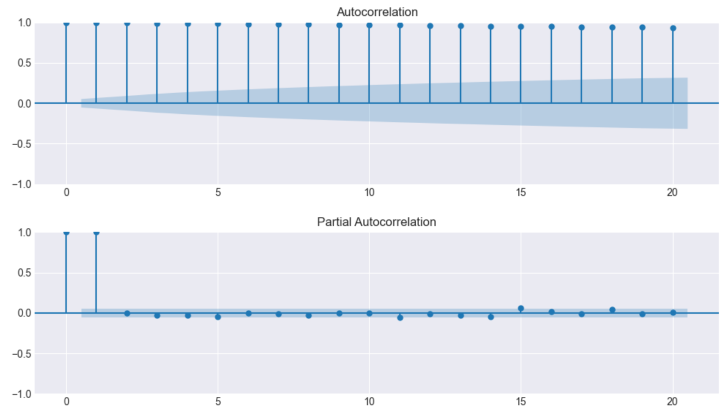
			JB Test		Shapiro Test		ADF ('c')			ADF ('ct')		
	Skew	Kurtosis	stats	p-value	stats	p-value	ADF-Stat	p-value	n-lags	ADF-Stat	p-value	**n-lags**
	Period 5											
BTCPrice	0,661	-0,701	18,495	0,000	0,906	0,000	-2,154	0,224	0	-3,705	0,022	2
Hashrate	-0,246	-1,102	12,108	0,002	0,951	0,000	-1,452	0,557	6	-3,038	0,122	4
Volume	1,388	1,753	86,292	0,000	0,864	0,000	-2,321	0,165	9	-3,086	0,109	7
Gold	-0,130	-1,543	20,080	0,000	0,900	0,000	-1,229	0,661	4	-0,920	0,954	4
Silver	-0,469	-1,229	19,698	0,000	0,888	0,000	-0,828	0,811	5	-2,361	0,400	2
Crudeoil	-0,135	-0,504	2,854	0,240	0,989	0,111	-1,926	0,320	0	-1,609	0,789	0
DXY	-0,422	-1,379	21,475	0,000	0,876	0,000	-0,874	0,796	0	-2,550	0,304	0
SP500	0,071	-0,870	6,563	0,038	0,976	0,001	-2,395	0,143	3	-3,179	0,089	3
Carbon	-0,323	-0,827	9,220	0,010	0,956	0,000	-1,912	0,326	0	-1,552	0,811	1
	Period 6											
BTCPrice	-0,107	-1,433	10,536	0,005	0,922	0,000	-1,454	0,556	0	-1,325	0,882	0
Hashrate	-0,430	-1,033	9,172	0,010	0,936	0,000	-1,353	0,605	2	-2,225	0,476	2
Volume	2,551	10,957	685,768	0,000	0,802	0,000	-4,273	0,000	1	-5,247	0,000	1
Gold	0,265	-0,616	3,467	0,177	0,970	0,007	-1,564	0,502	0	-1,656	0,770	0
Silver	-0,282	-0,163	1,795	0,408	0,954	0,000	-1,445	0,560	0	-1,639	0,777	0
Crudeoil	0,017	-1,161	6,881	0,032	0,957	0,001	-1,756	0,403	1	-2,096	0,548	1
DXY	-0,267	-1,217	8,939	0,011	0,937	0,000	-1,498	0,535	0	-1,650	0,772	0
SP500	-0,448	-0,675	6,441	0,040	0,948	0,000	-1,038	0,739	0	-3,410	0,050	0
Carbon	-0,638	-0,872	12,050	0,002	0,902	0,000	-1,710	0,426	0	-2,366	0,398	0
	Period 7											
BTCPrice	0,280	-1,250	9,389	0,009	0,931	0,000	-1,147	0,696	0	-2,510	0,323	0
Hashrate	-0,382	-0,220	3,220	0,200	0,984	0,164	-1,232	0,660	6	-11,018	0,000	0
Volume	1,104	0,830	26,796	0,000	0,910	0,000	-8,768	0,000	0	-9,510	0,000	0
Gold	1,308	1,351	41,510	0,000	0,879	0,000	-1,225	0,663	0	-2,355	0,404	7
Silver	0,536	-0,445	6,801	0,033	0,958	0,001	-2,959	0,039	5	-3,093	0,108	5
Crudeoil	0,781	0,815	14,687	0,001	0,958	0,001	-1,352	0,605	0	-1,704	0,749	0
DXY	0,352	-0,089	2,524	0,283	0,947	0,000	-1,126	0,704	0	-2,477	0,339	0
SP500	-0,294	-0,993	6,753	0,034	0,950	0,000	-1,798	0,381	0	-2,086	0,554	0
Carbon	-0,071	-1,433	10,313	0,006	0,920	0,000	-1,343	0,609	5	-1,638	0,777	5

or seasonality.

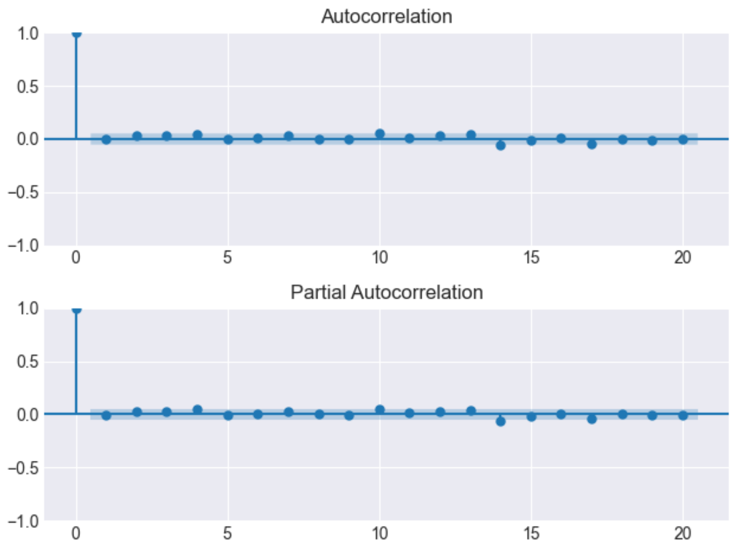
First, ACF and PACF plots are presented to visually estimate p and q values. Next, AIC values are calculated and compared for various model alternatives with different orders. Finally, model orders which minimize AIC value are selected for modeling ARIMA. Next, to validate an estimated model, the autocorrelation of residuals is checked by the Ljung-Box test and plotted for visual inspection. Finally, the characteristic roots of the models are examined to check whether the model is stationary and invertible. If the inverse roots from the characteristic polynomials are close to the center of the unit circle, then the roots are numerically stable, and the model is appropriate for forecasting. In case the model is not valid, the second candidate is chosen by AIC values.

According to [36, 32], a variable's statistical significance is typically not a good indicator in variable selection. For example, a variable helpful for forecasting might be associated with an insignificant coefficient. First, in case of a high correlation between two predictor variables, statistical tests of the coefficients may produce insignificant results because it is difficult to distinguish between each of their individual contributions, which causes a higher standard error on the coefficients. Second, an estimated coefficient may be significant in a large sample size despite a low coefficient and high variability. In this case, including the predictor may not be helpful while the variance of the forecast increases.

The autocorrelation function gives the correlation values between a time series and its lagged values. The partial autocorrelation shows the partial correlation coefficients between a series and lags of itself. Both plots can help identify AR and MA orders to fit an ARIMA model. By following this literature, $ARIMA$ model of BTC price in Period 1 is examined and validated as an example. The ACF plot shows the correlations with the lags are high and positive with a very slow decay in Figure 3.7a. At the same time, the PACF plot shows that the partial autocorrelations have a single spike at lag 1. The plots indicate the strong persistence in BTC price. After the differencing, the ACF and PACF plots are shown in Figure 3.7b. There is no significant spike in both plot. While lag 14 is outside the limits, lag 4 is on the limit in the PACF plot.



(a) Transformed BTC



(b) Differenced BTC

Figure 3.7: Period 1: ACF and PACF plots of BTC

AIC, BIC, and HQIC values of the different models with p and q orders from 1 to 5 are calculated, and the model with a minimum AIC value is selected for further analysis. The candidate models of BTC in Period 1 are presented in Table 3.9. $ARIMA(1, 1, 1)$ is chosen and the model coefficients are presented in Table 3.10. Ljung-Box test is applied to check

Table 3.9: Period 1: Candidate models for BTC

Model	AIC	BIC	HQIC
$ARIMA(1, 1, 1)$	**-4666.115**	**-4650.373**	**-4660.231**
$ARIMA(2, 1, 1)$	-4664.247	-4643.258	-4656.402
$ARIMA(1, 1, 2)$	-4663.346	-4642.358	-4655.502
$ARIMA(1, 1, 3)$	-4662.451	-4636.216	-4652.645
$ARIMA(2, 1, 2)$	-4662.155	-4635.919	-4652.349
$ARIMA(0, 1, 0)$	-4661.448	-4656.201	-4659.487
$ARIMA(4, 1, 3)$	-4660.890	-4618.913	-4645.200
$ARIMA(3, 1, 2)$	-4660.437	-4628.954	-4648.670
$ARIMA(2, 1, 3)$	-4660.433	-4628.950	-4648.666
$ARIMA(4, 1, 0)$	-4660.196	-4633.961	-4650.390

Table 3.10: Period 1: Model output of $ARIMA(1, 1, 1)$ for BTC

	Estimate	Std. Error	z	p-value
ar1	0.9750	0.018	53.128	0.000
ma1	-0.9576	0.024	-40.062	0.000
sigma2	0.0021	0.000	40.930	0.000

Log-likelihood	2336.057	AIC	-4666.115
BIC	-4650.373	HQIC	-4660.231

	z	p-value
Ljung-Box	0.16	0.69
Heteroskedasticity	0.96	0.66
Jarque-Bera (JB)	461.31	0.00

autocorrelation in the residuals of the model. According to the results in Table 3.10, the model satisfies the condition of independence in the residuals since the p-value is greater than the critical value 0.05, the null hypothesis that the residuals are independently distributed cannot be rejected. In addition, the Heteroskedasticity test also gives a p-value greater than 0.05 which indicates the residual distribution has constant variance. The JB test shows that the skewness and kurtosis values don't match a normal distribution. In Figure 3.8, the residuals are presented, and they have a mean around zero and a constant variance.

The summary of $ARIMA$ models are presented in Table 3.11. Autocorrelation plots, model outputs, and residual diagnostics can be found in Appendix. According to the results, BTC price follows a random walk process in Periods 3, 4, 6, and 7. In addition to BTC price, Gold is a random walk in Periods 1, 2, 4, and 7. While crude oil has random walk properties in Periods 3, 4, 5, and 6, DXY is in all periods except Period 3. $ARIMA(0, 1, 1)$ without intercept is equivalent to simple exponential smoothing, which is exhibited by hashrate in Period 2, carbon in Period 3, and crude oil in Period 5. On the other hand, $ARIMA(0, 1, 1)$ with intercept is equivalent to simple exponential smoothing with growth. This model fitted in hashrate in Periods 2 and 5, S&P500 in Period 3.

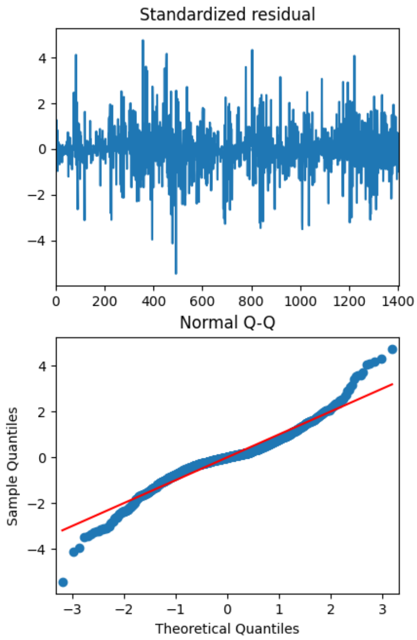


Figure 3.8: Period 1: Residuals of $ARIMA(1, 1, 1)$ model for BTC

Table 3.11: The orders of $ARIMA$ models

	Periods						
	1	2	3	4	5	6	7
BTCPrice	(1,1,1)	(1,1,0)*	(0,1,0)*	(0,1,0)	(3,0,0)*	(0,1,0)	(0,1,0)
Hashrate	(0,1,1)*	(0,1,1)	(2,0,2)*	(4,1,0)	(0,1,1)*	(2,1,0)	(3,0,2)*
Volume	(4,1,5)	(1,1,1)	(4,1,3)	(0,0,5)*	(0,1,2)	(1,0,1)	(1,0,0)*
Gold	(0,1,0)*	(0,1,0)	(1,1,0)	(0,1,0)*	(4,1,0)	(1,1,0)	(0,1,0)*
Silver	(0,1,2)	(0,1,2)	(0,1,0)	(3,1,2)	(0,1,3)	(0,1,0)	(2,2,2)
Crudeoil	(5,1,2)	(0,1,4)	(0,1,0)	(0,1,0)	(0,1,0)	(0,1,1)	(0,1,0)
DXY	(0,1,0)	(0,1,0)	(2,1,0)	(0,1,0)*	(0,1,0)	(0,1,0)	(0,1,0)*
SP500	(3,1,0)	(3,0,2)*	(0,1,1)*	(0,1,0)	(3,1,0)	(0,1,2)	(0,1,0)
Carbon	(1,1,1)*	(1,1,1)*	(0,1,1)	(0,1,0)	(2,1,2)	(0,1,0)*	(2,1,4)*

* models with an intercept

3.3 Smoothing Models

Exponential smoothing methods, forecasting moving averages, uses weighted averages of past observations in an exponential way in which the recent data takes the greater weight. There are smoothing parameters that are determined to give weights for previous observations. In this part, simple exponential smoothing, Holt's linear trend, and exponential trend are used and compared to forecast BTC price.

Simple Exponential Smoothing (SES), a method for data with no trend or seasonality, assumes that the last observation is the only one that contains valuable information for the future. The method's smoothing factor α, the only parameter, determines the rate at which the influence of the observations decays exponentially. First, l_0 and α values should be selected for the method. In here, l_0 equals to the first observation, on the other hand, α is an optimized value selected by Python's statsmodel.

Holt's trend method, an extension to SES, uses an additional parameter to determine the decay of the influence of the change in trend. Additive or multiplicative versions can be used to define the movement. While Holt's linear model is used for additivity, the exponential trend is used for multiplicativity in a trend. These parameters also chosen by Python's statsmodel. The coefficients of the models can be found in Appendix.

3.4 Forecast Results

Forecasts are evaluated by the accuracy metrics RMSE, MAE, and MAPE, and the results are shown in Tables 3.12, 3.13, 3.14, 3.15. In the flat periods 6 and 7, the forecasts have better performance compared to the up and down periods. The sub-periods are predicted with higher accuracy than the long period. The $ARIMA$ models give unstable performance in BTC price forecasting. While BTC price is well predicted by $ARIMA$ in Period 2, 4, especially in Periods 3 and 5, the accuracy is quite low. Gold, DXY and silver have stable accuracy values in all periods. Hashrate and BTC price exhibit similar performances in the sub-periods.

In Period 1, Gold and DXY exhibits a random walk process with $ARIMA(0, 1, 0)$, and the results $ARIMA$ and SES are very close. $ARIMA$ and SES models also outperform in the down market periods. In the flat periods 6 and 7, SES outperforms Holt's linear

Table 3.12: Period 1: Forecast Performance of $ARIMA$ and ES models

	$ARIMA$			SES			Holt's Linear			Exponential Trend		
Variable	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
BTC	**0.354**	**0.253**	**0.074**	0.564	0.462	0.075	0.634	0.582	0.109	0,650	0.603	0.119
Hashrate	0.134	0.103	0.005	0.288	0.252	0.013	0.112	0.096	0.005	**0.111**	**0.091**	**0.005**
Volume	**0.048**	**0.043**	**0.526**	0.048	0.044	1.525	0.055	0.049	1.760	0.043	0.033	0.682
Gold	**0.050**	**0.041**	**0.005**	0.051	0.041	0.005	0.071	0.053	0.007	0.072	0.053	0.007
Silver	**0.145**	**0.117**	**0.038**	0.145	0.117	0.039	0.187	0.147	0.049	0.185	0.146	0.048
Crude oil	0.170	0.138	0.030	0.170	0.138	0.030	0.138	0.118	0.026	**0.138**	**0.117**	**0.026**
Dollar index	**0.103**	**0.086**	**0.018**	0.103	0.086	0.025	0.110	0.092	0.019	0.108	0.900	0.019
SP500	0.214	0.176	0.021	**0.125**	**0.100**	**0.012**	0.217	0.178	0.021	0.222	0.183	0.022
Carbon	0.201	0.176	0.040	0.341	0.323	0.073	0.201	0.176	0.040	**0.211**	**0.170**	**0.039**

and exponential trend models. While $ARIMA$ gives more accurate results for BTC price in Period 1, 6 and 7, Holt's linear and exponential trend models have better performance in Periods 2 and 3 respectively.

Multivariate time-series models may give better predictions compared to $ARIMA$ and other univariate models since multivariate models take into account additional variables which may have a predictive power between each other. In Table 3.16, the forecast performance is presented for VAR model of BTC and hashrate. For BTC price, there is no significant im-

Table 3.13: Forecast Performance of ARIMA Models

	Period 2			Period 3			Period 4		
Variable	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
BTC	0.183	0.147	0.013	0.459	0.408	0.042	0.123	0.106	0.010
Hashrate	0.190	0.171	0.009	0.404	0.364	0.022	0.097	0.080	0.004
Volume	0.138	0.103	0.448	0.233	0.145	0.361	0.022	0.021	0.352
Gold	0.024	0.021	0.002	0.025	0.018	0.002	0.014	0.011	0.002
Silver	0.039	0.033	0.010	0.038	0.029	0.010	0.044	0.036	0.012
Crude oil	0.118	0.102	0.025	0.007	0.059	0.014	0.036	0.030	0.007
Dollar index	0.019	0.017	0.003	0.006	0.005	0.001	0.029	0.025	0.006
SP500	0.048	0.042	0.005	0.029	0.025	0.003	0.051	0.046	0.005
Carbon	0.052	0.043	0.012	0.037	0.030	0.014	0.120	0.098	0.022
Variable	Period 5			Period 6			Period 7		
BTCPrice	0.563	0.505	0.061	0.068	0.061	0.006	0.079	0.063	0.006
Hashrate	0.410	0.382	0.021	0.064	0.055	0.021	0.183	0.146	0.008
Volume	0.077	0.055	0.450	0.057	0.054	0.060	0.033	0.025	0.461
Gold	0.016	0.013	0.002	0.007	0.006	0.001	0.099	0.097	0.013
Silver	0.024	0.019	0.007	0.027	0.024	0.007	0.020	0.017	0.005
Crudeoil	0.205	0.184	0.047	0.046	0.040	0.010	0.093	0.108	0.019
DXY	0.007	0.006	0.001	0.006	0.005	0.001	0.060	0.059	0.013
SP500	0.060	0.045	0.006	0.012	0.010	0.001	0.034	0.032	0.004
Carbon	0.213	0.067	0.233	0.034	0.029	0.007	0.206	0.168	0.040

provement in the forecast of VAR model compared to $ARIMA$. While $ARIMA$ model has the values of RMSE (0.183), MAE (0.147), and MAPE (0.013), VAR model gives slightly better results. Figure 3.9 shows the forecast of the models.

Table 3.14: Forecast Performance of Exponential Smoothing Models: Periods 2-4

Period 2									
	Simple Exponential			Holt's Linear			Exponential Trend		
Variable	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
BTC	0.317	0.275	0.026	**0.141**	**0.114**	**0.011**	0,147	0,122	0,012
Hashrate	0.102	0.089	0.005	**0.088**	**0.072**	**0.004**	0.088	0.072	0.004
Volume	0.135	0.099	0.429	0.131	0.093	0.403	**0.128**	**0.090**	**0.390**
Gold	**0.025**	**0.021**	**0.003**	0.038	0.034	0.005	0.038	0.034	0.004
Silver	**0.039**	**0.033**	**0.010**	0.071	0.066	0.020	0.071	0.066	0.020
Crude oil	0.128	0.123	0.028	**0.047**	**0.040**	**0.010**	0.051	0.043	0.011
Dollar index	**0.010**	**0.009**	**0.002**	0.019	0.017	0.004	0.019	0.017	0.003
SP500	0.031	0.027	0.003	**0.015**	**0.011**	**0.001**	0.015	0.012	0.001
Carbon	0.107	0.085	0.024	0.053	0.044	0.012	**0.052**	**0.043**	**0.012**
Period 3									
Variable	Simple Exponential			Holt's Linear			Exponential Trend		
BTCPrice	0.648	0.589	0.061	0.460	0.409	0.043	**0.429**	**0.378**	**0.039**
Hashrate	0.347	0.312	0.019	0.165	0.141	0.009	**0.157**	**0.133**	**0.008**
Volume	0.241	0.155	0.390	**0.10**	**0.124**	**0.338**	0.214	0.144	0.482
Gold	**0.019**	**0.016**	**0.002**	0.020	0.016	0.002	0.020	0.016	0.002
Silver	**0.038**	**0.030**	**0.011**	0.041	0.031	0.011	0.041	0.031	0.011
Crudeoil	0.071	0.059	0.015	**0.067**	**0.056**	**0.014**	0.068	0.057	0.014
DXY	0.012	0.009	0.002	**0.006**	**0.005**	**0.001**	0.006	0.005	0.001
SP500	0.043	0.038	0.005	**0.029**	**0.025**	**0.003**	0.029	0.026	0.003
Carbon	**0.037**	**0.031**	**0.015**	0.043	0.035	0.017	0.070	0.060	0.029
Period 4									
Variable	Simple Exponential			Holt's Linear			Exponential Trend		
BTCPrice	**0.054**	**0.043**	**0.004**	0.124	0.107	0.011	0,124	0,106	0,107
Hashrate	**0.094**	**0.076**	**0.004**	0.123	0.090	0.005	0.124	0.090	0.005
Volume	**0.012**	**0.008**	**0.345**	0.014	0.009	0.362	0.018	0.015	0.603
Gold	**0.012**	**0.010**	**0.001**	0.015	0.011	0.001	0.015	0.011	0.001
Silver	**0.043**	**0.033**	**0.011**	0.057	0.043	0.014	0.057	0.043	0,015
Crudeoil	0.037	0.030	0.007	0.035	0.029	0.007	**0.035**	**0.030**	**0.006**
DXY	**0.017**	**0.014**	**0.003**	0.029	0.025	0.005	0.029	0.005	0.005
SP500	**0.052**	**0.046**	**0.006**	0.071	0.064	0.008	0.071	0.064	0.008
Carbon	**0.119**	**0.098**	**0.022**	0.137	0.115	0.026	0.138	0.115	0.027

Table 3.15: Forecast Performance of Exponential Smoothing Models: Periods 5-7

Period 5									
	Simple Exponential			Holt's Linear			Exponential Trend		
Variable	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
BTC	0.474	0.422	0.051	0.423	0.376	0.046	**0.422**	**0.375**	**0.046**
Hashrate	**0.276**	**0.250**	**0.014**	0.411	0.383	0.022	0.413	0.385	0.022
Volume	**0.086**	**0.067**	**0.581**	0.100	0.084	0.794	0.097	0.079	0.737
Gold	**0.016**	**0.013**	**0.002**	0.021	0.017	0.002	0.021	0.017	0.002
Silver	**0.025**	**0.019**	**0.007**	0.035	0.027	0.010	0.035	0.027	0.010
Crude oil	0.205	0.184	0.047	0.203	0.182	0.047	**0.203**	**0.181**	**0.047**
Dollar index	0.007	0.006	0.001	0.005	0.004	0.001	**0.005**	**0.004**	**0.001**
SP500	0.060	0.046	0.006	**0.057**	**0.043**	**0.005**	0.057	0.043	0.005
Carbon	0.196	0.174	0.054	**0.121**	**0.106**	**0.033**	0.124	0.109	0.034
Period 6									
Variable	Simple Exponential			Holt's Linear			Exponential Trend		
BTCPrice	**0.067**	**0.059**	**0.006**	0.085	0.079	0.007	0.087	0.081	0.007
Hashrate	**0.104**	**0.080**	**0.004**	0.127	0.104	0.005	0.127	0.104	0.006
Volume	0.030	0.017	0.290	**0.033**	**0.017**	**0.245**	0.034	0.018	0.252
Gold	0.007	0.006	0.001	**0.007**	**0.006**	**0.001**	0.007	0.006	0.001
Silver	**0.022**	**0.017**	**0.005**	0.027	0.024	0.008	0.027	0.024	0.008
Crudeoil	0.046	0.040	0.010	**0.037**	**0.031**	**0.007**	0.037	0.031	0.007
DXY	**0.006**	**0.005**	**0.001**	0.010	0.009	0.002	0.010	0.009	0.002
SP500	0.020	0.009	0.001	**0.011**	**0.009**	**0.001**	0.011	0.009	0.001
Carbon	0.054	0.047	0.012	**0.034**	**0.029**	**0.007**	0.035	0.029	0.007
Period 7									
Variable	Simple Exponential			Holt's Linear			Exponential Trend		
BTCPrice	**0.080**	**0.064**	**0.006**	0.082	0.066	0.006	0.083	0.066	0.006
Hashrate	**0.053**	**0.043**	**0.002**	0.095	0.080	0.004	0.095	0.080	0.004
Volume	0.025	0.021	0.550	**0.022**	**0.016**	**0.397**	0.022	0.017	0.438
Gold	0.012	0.010	0.001	**0.008**	**0.007**	**0.001**	0.008	0.007	0.001
Silver	0.025	0.019	0.006	**0.020**	**0.018**	**0.005**	0.020	0.018	0.005
Crudeoil	0.108	0.093	0.020	0.093	0.076	0.016	**0.092**	**0.076**	**0.016**
DXY	0.008	0.007	0.001	0.005	0.005	0.001	**0.005**	**0.005**	**0.001**
SP500	**0.035**	**0.032**	**0.004**	0.035	0.032	0.004	0.035	0.032	0.004
Carbon	**0.019**	**0.015**	**0.003**	0.026	0.023	0.005	0.025	0.021	0.005

Table 3.16: Forecast Performance of VAR Model BTC&Hashrate

Variable	RMSE	MAE	MAPE
BTC	0.174	0,139	0,013
Hashrate	0.082	0.064	0.003

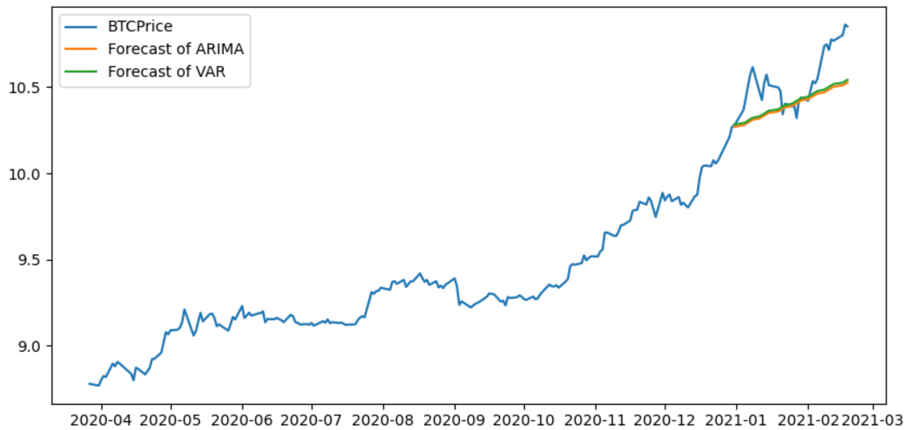


Figure 3.9: Period 2: Forecasts of BTC price

On the other hand, according to the granger-causality test in Period 2, there is a unidirectional granger-causality from BTC price to hashrate, indicating the predictive power of BTC price on hashrate. Since this relationship can improve the accuracy of the forecast for hashrate in Period 2, the forecasting ability of VAR model is presented in Table 3.16, and compared with $ARIMA$ model. While $ARIMA(0, 1, 1)$ model of hashrate gives the values of RMSE (0.190), MAE (0.171), VAR model have better performance with RMSE (0.082), MAE (0.064) values. Thus, the results indicate that BTC price is helpful in forecasting hashrate.

3.5 ECM-VAR Models

In time series, vector autoregressive (VAR) models are frequently used to search for the dynamic relationships between variables. The design of VAR models enables the examination of endogenous variable values from their observed historical values. Engle and Granger's [23] contributions also offered methods for modeling cointegrated relationships between variables. Before making the variables stationary for VAR model, it is possible to investigate the cointegration relationship between non-stationary variables.

3.5.1 Cointegration

Suppose that two non-stationary time series have the same order of integration, and there may be a stationary linear combination of these series. If a cointegration is found, an error correction model (ECM) can be estimated for the stochastic trend and equilibrium. In this study, the Engle-Granger two-step method [23] is applied for the cointegration test. Engle-Granger's two-step method has some specific steps that should be followed, as below.

1. Stationary testing for the time series. If the series are non-stationary in level and their order of integration values is equal, then the next step is followed.

2. A linear model is built for the series by OLS, and the residuals of the linear model are checked for stationarity by the Augmented Dickey-Fuller test. If the residuals are stationary, cointegration is found. Conversely, the non-stationarity of the residuals indicates the absence of cointegration.

The cointegration analysis cannot be applied, in case a time series is trend-stationary. Since BTC price is trend-stationary in Period 5, and hashrate is in Periods 3 and 7, Periods 2, 4, and 6 are selected to investigate the cointegration relationships of the variables.

In Table 3.17, ADF test results are shown for the residuals of the linear models. Since S&P500 is trend stationary in Period 2, and volume is stationary in Periods 4, 6, these variables are not included in the cointegration studies. We should use the critical values from [23] to evaluate the test statistics.

Table 3.17: ADF test results for the residuals of the regressions

	Period 2	Period 4	Period 6
	Test statistics		
BTCHashrate	-4,8600	-3,1086	-2,8005
BTCVolume	-3,0899	-	-
BTCGold	1,4174	-2,2994	-1,9742
BTCSilver	1,3958	-1,7451	-1,4748
BTCCrudeoil	0,1198	-3,1883	-3,0458
BTCDXY	0,4748	-1,6771	-1,3196
BTCSP500	-	-1,4779	-1,6440
BTCCarbon	-0,2972	-1,8042	-1,7124

By following the two-step method, the linear equations for BTC and hashrate is as below. The stationarity of the error term u_t is tested by ADF test. Since the regression has constant β_0 in Table 3.18, ADF test is applied without a constant. Only the residuals of BTC and hashrate have lower statistics than the critical values in Period 2. Hence, the null hypothesis of the ADF test can be rejected. The other equations have higher values than the test's critical values, which indicates the absence of cointegration.

$$
\begin{aligned}
Hash_t &= \beta_0 + \beta_1 BTC_t + u_t \\
Hash_t &= 16.7695 + 0.1973 * BTC_t + u_t
\end{aligned}
$$

Table 3.18: Regression of hashrate and BTC

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	16.7695	0.2581	64.97	0.0001
BTC	0.1973	0.2764	7.14	0.0001

To investigate the cointegration relationship between BTC and hashrate, the lag of the error correction model should be determined first. A VAR model is estimated with these non-stationary variables to specify the lag order which gives the minimum AIC value. The error correction model is estimated with a lag of 1, and the equations of the model are shown in Table 3.19. If there is a negative and significant lagged error correction term (ECT) which is an adjustment coefficient, BTC and hashrate can have a common stochastic trend and reach the equilibrium. In the case of equilibrium, the effect of a shock is not permanent in the long-run relationship. In the equation of BTC as an independent variable, the error correction term is small and insignificant. On the other hand, the equation of hashrate has a significant negative ECT value of -0.1647, which indicates the speed of adjustment towards the equilibrium in $1/0.1647$ periods. the common stochastic trend of BTC and hashrate. This also implies a granger-causality relationship from BTC to hashrate. The equations are also presented in matrix form below.

$$\begin{aligned} BTC.d_t &= c_1 + \alpha_1 * ECT_{t-1} + a_{1,1}BTC.d_{t-1} + a_{1,2}Hash.d_{t-1} + e_{1,t} \\ Hash.d_t &= c_2 + \alpha_2 * ECT_{t-1} + a_{2,1}BTC.d_{t-1} + a_{2,2}Hash.d_{t-1} + e_{2,t} \end{aligned}$$

$$\begin{bmatrix} BTC.d_t \\ Hash.d_t \end{bmatrix} = \begin{bmatrix} -0.0002 \\ 0.0001 \end{bmatrix} + \begin{bmatrix} 0.0058 \\ -0.1647 \end{bmatrix} \begin{bmatrix} ECT_{t-1} \\ ECT_{t-1} \end{bmatrix} + \begin{bmatrix} -0.1097 & 0.0030 \\ -0.4508 & 0.4389 \end{bmatrix} \begin{bmatrix} BTC.d_{t-1} \\ Hash.d_{t-1} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix}$$

$$\begin{aligned} BTC_t &= -0.0002 + 0.0058 \times ECT_1 - 0.1097 * BTC.d_{t-1} + 0.0030 \times Hash.d_{t-1} \\ Hash_t &= -0.0001 - 0.1647 \times ECT_2 - 0.4508 * BTC.d_{t-1} + 0.4389 \times Hash.d_{t-1} \end{aligned}$$

Table 3.19: Error correction model of BTC and hashrate

	Estimate	Std. Error	t value	Pr(>ltl)
Intercept	-0.0001	0.0011	-0.115	0.908
ECT	-0.1647	0.0189	-8.697	0.001
Hashrate.d1	0.4389	0.0366	11.985	0.001
BTC.d1	-0.4508	0.2234	0.737	0.461

(a) Estimated equation for hashrate

	Estimate	Std. Error	t value	Pr(>ltl)
Intercept	-0.0002	0.0002	-1.357	0.1755
ECT	0.0058	0.0036	1.618	0.1064
Hashrate.d1	0.0030	0.0070	0.432	0.0109
BTC.d1	-0.1097	0.0429	-2.554	0.6656

(b) Estimated equation for BTC

3.5.2 VAR

Since S&P500 is trend stationary in Period 2, the variable is detrended before estimating a VAR model. Also, volume is stationary in Periods 4 and 6, no differencing is applied to volume. In Table 3.20, the lags of VAR models, determined by AIC, are shown for Periods 2, 4, and 6. Accordingly, to these lag values, VAR models are estimated for each pair of BTC. The serial correlation of the residuals is checked by Breusch–Godfrey (BG) test, and no serial correlation is found in the residuals. Table 3.21 shows the results of BG test. The residuals of the VAR models can be seen in Figures 3.10a, 3.10b, 3.10c. The residuals have mean zero and constant variance.

Table 3.20: Lags of VAR models

	Period 2	Period 4	Period 6
BTCHashrate	2	4	2
BTCVolume	5	1	2
BTCGold	1	1	1
BTCSilver	1	1	1
BTCCrudeoil	4	1	1
BTCDXY	1	1	1
BTCSP500	3	1	1
BTCCarbon	1	1	1

Three VAR models in which granger causality is found are presented in Table 3.22. The matrix forms of the models are below. The stability of coefficients in a multivariate model can be ensured by CUSUM test. In Figures 3.11a, 3.11b, 3.11c, CUSUM test results are shown for the residuals of VAR models. The test shows that there is no structural break in the residuals.

Table 3.21: Breusch–Godfrey test results of the residuals

	Period 2	Period 4	Period 6
	p-values		
BTCHashrate	0.9588	0.5961	0.7431
BTCVolume	0.3813	0.1614	0.1863
BTCGold	0.2507	0.7737	0.5873
BTCSilver	0.5520	0.8060	0.7615
BTCCrudeoil	0.1786	0.5756	0.3355
BTCDXY	0.2572	0.6670	0.4621
BTCSP500	0.1262	0.7957	0.7553
BTCCarbon	0.3264	0.4798	0.8895

In the VAR model of BTC and hashrate in Period 2, the lag of BTC has a significant positive value of 0.4051 which indicates the explanatory power of BTC price in hashrate. However, there is no significant value of hashrate in the equation of BTC price.

Period 2: BTC&Hashrate

$$\begin{bmatrix} BTC_t \\ Hash_t \end{bmatrix} = \begin{bmatrix} 0.0094 \\ -0.0004 \end{bmatrix} + \begin{bmatrix} -0.1581 & -0.0288 \\ 0.4051 & -0.6140 \end{bmatrix} \begin{bmatrix} BTC_{t-1} \\ Hash_{t-1} \end{bmatrix} + \begin{bmatrix} -0.0113 & 0.0519 \\ 0.0949 & -0.3168 \end{bmatrix} \begin{bmatrix} BTC_{t-2} \\ Hash_{t-2} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix}$$

The VAR model of BTC and volume in Period 2 is estimated with a lag of 3. In the equation of

BTC, there is only one significant coefficient which is the first lag of BTC price. On the other hand, the equation of volume has there significant coefficients of BTC price with the lags of 1, 2 and 4. The coefficients are 0.2470, 0.2332, and 0.1118 which indicates the predictive power of BTC price on volume.

Period 2: BTC&Volume

$$\begin{bmatrix} BTC_t \\ Vol_t \end{bmatrix} = \begin{bmatrix} 0.0006 \\ 0.0071 \end{bmatrix} + \begin{bmatrix} -0.1750 & 0.0650 \\ 0.2470 & 0.2814 \end{bmatrix} \begin{bmatrix} BTC_{t-1} \\ Vol_{t-1} \end{bmatrix} + \begin{bmatrix} -0.0114 & 0.0843 \\ 0.2332 & 0.1300 \end{bmatrix} \begin{bmatrix} BTC_{t-2} \\ Vol_{t-2} \end{bmatrix}$$
$$+ \begin{bmatrix} 0.0681 & 0.1251 \\ 0.0245 & -0.0621 \end{bmatrix} \begin{bmatrix} BTC_{t-3} \\ Vol_{t-3} \end{bmatrix} + \begin{bmatrix} -0.0664 & -0.0268 \\ 0.1118 & 0.1018 \end{bmatrix} \begin{bmatrix} BTC_{t-4} \\ Vol_{t-4} \end{bmatrix}$$
$$+ \begin{bmatrix} -0.0762 & -0.0110 \\ 0.0866 & 0.2623 \end{bmatrix} \begin{bmatrix} BTC_{t-5} \\ Vol_{t-5} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix}$$

The VAR model of BTC and volume in Period 4 is estimated with one lag. In the equation of BTC, there is no significant coefficient. However, the equation of volume has there significant coefficients, including the first lag of BTC price.

Period 4: BTC&Volume

$$\begin{bmatrix} BTC_t \\ Vol_t \end{bmatrix} = \begin{bmatrix} -0.0062 \\ 0.0173 \end{bmatrix} + \begin{bmatrix} -0.0522 & -0.0175 \\ -0.2065 & 0.5623 \end{bmatrix} \begin{bmatrix} BTC_{t-1} \\ Vol_{t-1} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix}$$

3.5.3 Granger Causality

Granger causality determines whether a time series is useful in forecasting another. If a time series has statistically significant information about future values of another time series, then the time series granger causes the other one. The test can be done through a F test or Wald test. In the thesis, F-test is applied for granger-causality in the mean on the estimated coefficients of the bivariate VAR models.

The null hypothesis of the Granger causality test indicates the absence of a causality relationship. The results of the F-test, applied to VAR models, are shown in Table 3.23. In Period 2, two unidirectional granger-causality are found. According to the results, BTC price granger-causes hashrate and volume. Additionally, in Period 4, BTC price also granger-causes volume. On the other hand, there is no evidence of granger-causality in Period 6. While BTC price only granger-causes hashrate in the upward market, volume is granger-caused by BTC price in both upward and downward market conditions.

Table 3.22: VAR Models

	BTC				Hashrate			
	Estimate	Std. Error	t value	p value	Estimate	Std. Error	t value	p value
const	0.0094	0.0029	3.214	0.0015	-0.0004	0.0075	-0.054	0.9574
BTC.L1	-0.1581	0.0754	-2.096	0.0374	0.4051	0.1988	2.038	0.0430
Hash.L1	0.0288	0.0280	1.029	0.3046	-0.6140	0.0718	-8.546	0.0001
BTC.L2	-0.0113	0.0788	-0.143	0.8860	0.0949	0.2020	0.470	0.6387
Hash.L2	0.0519	0.0280	1.856	0.0651	-0.3168	0.0717	-4.414	0.0001

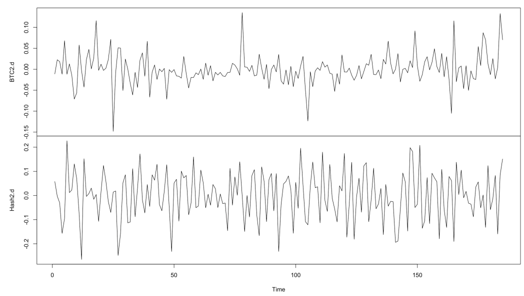
a) Period 2: BTC and hashrate

	BTC				Volume			
	Estimate	Std. Error	t value	p value	Estimate	Std. Error	t value	p value
const	0.0006	0.0058	0.110	0.9125	0.0071	0.0034	2.088	0.0383
BTC.L1	-0.1750	0.0776	-2.253	0.0255	0.2470	0.0453	5.451	0.0001
Vol.L1	0.0650	0.1272	0.511	0.6100	0.2814	0.0742	3.793	0.0002
BTC.L2	-0.0114	0.0873	-0.131	0.8961	0.2332	0.0509	4.578	0.0001
Vol.L2	0.0843	0.1321	0.638	0.5242	0.1300	0.0770	1.688	0.0933
BTC.L3	0.0681	0.0915	0.744	0.4577	0.0245	0.0534	0.460	0.6462
Vol.L3	0.1251	0.1336	0.937	0.3503	-0.0621	0.0779	-0.797	0.4266
BTC.L4	-0.0664	0.0925	-0.718	0.4735	0.1118	0.0539	2.072	0.0397
Vol.L4	-0.0268	0.1286	-0.209	0.8349	0.1018	0.0750	1.357	0.1765
BTC.L5	-0.0762	0.0905	-0.843	0.4007	0.0866	0.0527	1.641	0.1026
Vol.L5	-0.0110	0.1232	-0.090	0.9285	0.2623	0.0718	3.651	0.0003

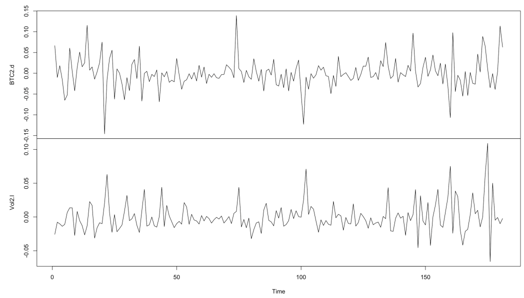
b) Period 2: BTC and volume

	BTC				Volume			
	Estimate	Std. Error	t value	p value	Estimate	Std. Error	t value	p value
const	-0.0062	0.0071	-0.879	0.381	0.0173	0.0039	4.359	0.0001
BTC.L1	-0.0522	0.0936	-0.558	0.578	-0.2065	0.0521	-3.959	0.0001
Vol.L1	-0.0175	0.1299	-0.135	0.893	0.5623	0.0724	7.766	0.0001

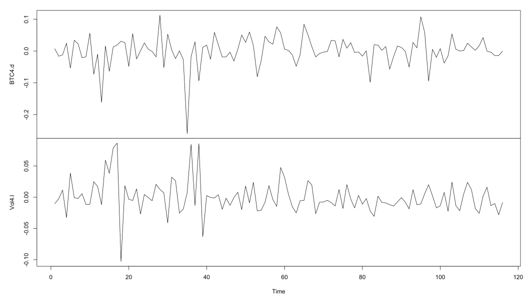
c) Period 4: BTC and volume



(a) Period 2: BTC&Hashrate

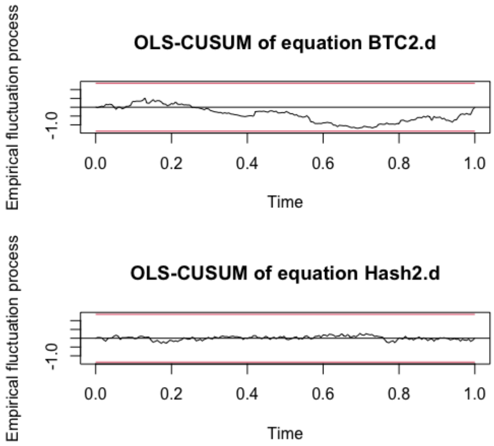


(b) Period 2: BTC&Volume

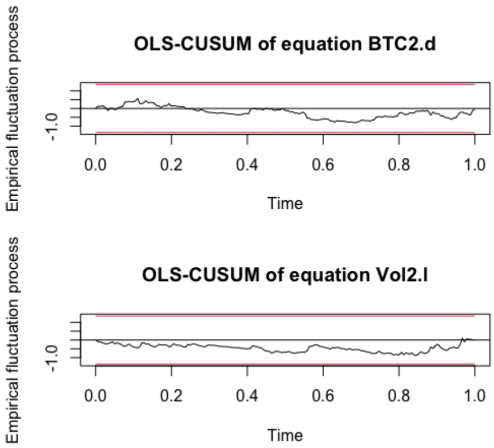


(c) Period 4: BTC&Volume

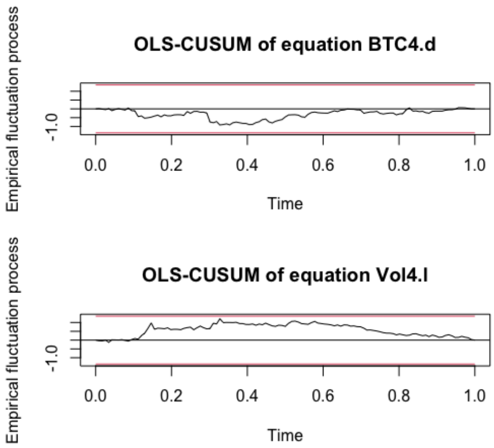
Figure 3.10: Residuals of VAR models



(a) Period 2: BTC&Hashrate



(b) Period 2: BTC&Volume



(c) Period 4: BTC&Volume

Figure 3.11: OLS-CUSUM of VAR models

Table 3.23: Granger causality test results

	F-test	p-value		F-test	p-value
BTC > Hashrate	4.1517	**0.0420**	Hashrate > BTC	0.0160	0.8993
BTC > Volume	10.763	**0.0001**	Volume > BTC	0.1587	0.9589
BTC > Gold	2.5470	0.0564	Gold > BTC	0.0709	0.7901
BTC > Silver	3.1757	0.0755	Silver > BTC	1.2920	02564
BTC > Crudeoil	1.9665	0.1197	Crudeoil > BTC	1.4820	0.2195
BTC > DXY	3.8186	0.0514	DXY > BTC	0.0064	0.9363
BTC > SP500	1.9599	0.0840	SP500 > BTC	1.5970	0.1247
BTC > Carbon	0.0452	0.8317	Carbon > BTC	2.1881	0.1399

a) Period 2

	F-test	p-value		F-test	p-value
BTC > Hashrate	1.2991	0.2715	Hashrate > BTC	0.3534	0.8414
BTC > Volume	15.6730	**0.0001**	Volume > BTC	0.0182	0.8929
BTC > Gold	2.9224	0.0887	Gold > BTC	0.2647	0.6074
BTC > Silver	1.5983	0.1903	Silver > BTC	0.0085	0.9265
BTC > Crudeoil	1.8813	0.1545	Crudeoil > BTC	0.1829	0.8926
BTC > DXY	0.8164	0.3672	DXY > BTC	0.0435	0.8349
BTC > SP500	0.3241	0.5697	SP500 > BTC	0.9048	0.3425
BTC > Carbon	0.5700	0.4510	Carbon > BTC	1.4016	0.2243

b) Period 4

	F-test	p-value		F-test	p-value
BTC > Hashrate	1.2812	0.2797	Hashrate > BTC	1.0678	0.3455
BTC > Volume	1.0864	0.3390	Volume > BTC	1.8171	0.1648
BTC > Gold	0.1352	0.7133	Gold > BTC	0.0332	0.8555
BTC > Silver	0.1595	0.6900	Silver > BTC	1.2408	0.2665
BTC > Crudeoil	1.5051	0.2211	Crudeoil > BTC	0.0441	0.8338
BTC > DXY	0.0894	0.7652	DXY > BTC	1.8569	0.1743
BTC > SP500	0.2990	0.7418	SP500 > BTC	1.8813	0.1545
BTC > Carbon	0.0156	0.9005	Carbon > BTC	0.4634	0.4967

c) Period 6

CHAPTER 4

CONCLUSION AND DISCUSSION

The behavior of BTC price and its predictability attract the world of finance and the academic community. Therefore, the motivation for discovering an indicator for BTC price is a lasting topic in the literature. Thus, the primary goal of this thesis is to investigate the potential drivers of BTC prices. For this purpose, various variables, including network metrics such as the computational power on the network, also the commodities which have common characteristics with BTC, are included in the study.

Due to the unstable nature of BTC, the likelihood of shifting relationships is considered, and three periods with different market conditions are chosen by the practice of Prophet's trend changepoints. These periods are separately examined to answer the research question of how these relationships differ in different time horizons. The first period includes the bull period beginning around the last halving event in March 2020. While the second involves the ongoing dramatic downfall of BTC, starting from April 2022, the last period includes the flat movement in 2021.

The research question, "Does Bitcoin have any causal relationship with the variables?", is answered using multivariate models in three periods. In the first period, the cointegration relationship is found between BTC and hashrate which indicates a common stochastic trend. The cointegration reveals a unidirectional granger-causality from BTC price to hashrate in the upward market. The findings consistent with studies [24, 29] support the mining equilibrium of BTC network in which the increasing profit led by the rise of BTC price is an incentive for the miners to participate in mining and allocate more computational power.

The variables' predictive power is examined by Granger test, and the granger-causality is found between BTC price and volume. BTC price granger causes the trading volume in Periods 2 and 4. This finding highlights that the movements of BTC price can trigger changes in its market activity via trading volume. In other words, the trading volume is driven by BTC prices in both up and down market conditions. However, neither cointegration nor granger causality is found in the flat market condition.

Forecasting BTC price is another objective of this thesis. BTC price and other variables are predicted by $ARIMA$ and ES models in seven different periods, and the performance of the forecasts is compared to answer the research question of how accurately $ARIMA$ and ES models predict BTC price in various time periods. The forecasts in the sub-periods show higher accuracy than the whole period. The Dollar index and gold have better accuracy results in all periods. The forecasts in flat market conditions have lower errors than in upward and downward markets. The forecast performance of BTC price highly depends on the time intervals and varies a lot in the sub-periods. While BTC is accurately predicted in Periods 2 and 4, the performances in Periods 3 and 5 have the largest errors among the other variables.

Although the selected variables have no predictive ability on BTC price, the predictive power of BTC price is examined on hashrate by comparing the forecast performances of $ARIMA$ and VAR models. VAR model outperforms $ARIMA$ model for hashrate, and the results show that BTC price is a useful indicator in forecasting hashrate. On the other hand, VAR model does not provide a significant improvement for the forecast of BTC price, indicating that the lagged values of hashrate do not contain valuable information about BTC price.

In further studies, the cointegration relationship between hashrate and BTC can be investigated from an economic perspective. The potential mining equilibrium may help answer the questions that arise with the huge energy consumption of BTC network. The predictive power of BTC price on hashrate can be used as an indicator to forecast the energy consumption of BTC network. Finally, the seasonality properties of the halving cycle of BTC can be studied before the subsequent one in 2024.

REFERENCES

[1] H. Akaike, A new look at the statistical model identification, IEEE transactions on automatic control, 19(6), pp. 716–723, 1974.

[2] C. G. Akcora, A. K. Dey, Y. R. Gel, and M. Kantarcioglu, Forecasting bitcoin price with graph chainlets, in *Pacific-Asia conference on knowledge discovery and data mining*, pp. 765–776, Springer, 2018.

[3] A. Azari, Bitcoin price prediction: An arima approach, arXiv preprint arXiv:1904.05315, 2019.

[4] L. Badea and M. C. Mungiu-Pup$\tilde{a}$zan, The economic and environmental impact of bitcoin, IEEE Access, 9, pp. 48091–48104, 2021.

[5] J. Bai and P. Perron, Estimating and testing linear models with multiple structural changes, Econometrica, pp. 47–78, 1998.

[6] M. Balcilar, E. Bouri, R. Gupta, and D. Roubaud, Can volume predict bitcoin returns and volatility? a quantiles-based approach, Economic Modelling, 64, pp. 74–81, 2017.

[7] S. A. Basher and P. Sadorsky, Forecasting bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility?, Machine Learning with Applications, p. 100355, 2022.

[8] J. Bouoiyour and R. Selmi, Bitcoin price: Is it really that new round of volatility can be on way?, MPRA Paper, 2015.

[9] J. Bouoiyour and R. Selmi, What does bitcoin look like?, Annals of Economics & Finance, 16(2), 2015.

[10] E. Bouri, R. Gupta, C. K. M. Lau, D. Roubaud, and S. Wang, Bitcoin and global financial stress: A copula-based approach to dependence and causality in the quantiles, The Quarterly Review of Economics and Finance, 69, pp. 297–307, 2018.

[11] G. E. Box and D. R. Cox, An analysis of transformations, Journal of the Royal Statistical Society: Series B (Methodological), 26(2), pp. 211–243, 1964.

[12] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*, John Wiley & Sons, 2015.

[13] R. G. Brown, Exponential smoothing for predicting demand, in *Operations Research*, volume 5, pp. 145–145, Inst Operations Research Management Sciences 901 ELKRIDGE LANDING RD, STE . . . , 1957.

[14] R. Caferra and D. Vidal-Tomás, Who raised from the abyss? a comparison between cryptocurrency and stock market dynamics during the covid-19 pandemic, Finance Research Letters, 43, p. 101954, 2021.

[15] G. M. Caporale and T. Zekokh, Modelling volatility of cryptocurrencies using markov-switching garch models, Research in International Business and Finance, 48, pp. 143–155, 2019.

[16] Z. Chen, C. Li, and W. Sun, Bitcoin price prediction using machine learning: An approach to sample dimension engineering, Journal of Computational and Applied Mathematics, 365, p. 112395, 2020.

[17] U. W. Chohan, A history of bitcoin, Available at SSRN 3047875, 2022.

[18] S. Corbet, A. Meegan, C. Larkin, B. Lucey, and L. Yarovaya, Exploring the dynamic relationships between cryptocurrencies and other financial assets, Economics Letters, 165, pp. 28–34, 2018.

[19] N. B. da Silva Rosa, *Virtual currency: a cointegration analysis between bitcoin prices and economic and financial data*, Ph.D. thesis, University Institute of Lisbon, Av. das Forças Armadas, 1649-026 Lisboa, Portugal, 9 2017.

[20] S. Dastgir, E. Demir, G. Downing, G. Gozgor, and C. K. M. Lau, The causal relationship between bitcoin attention and bitcoin returns: Evidence from the copula-based granger causality test, Finance Research Letters, 28, pp. 160–164, 2019.

[21] D. A. Dickey and W. A. Fuller, Distribution of the estimators for autoregressive time series with a unit root, Journal of the American statistical association, 74(366a), pp. 427–431, 1979.

[22] L. dos Santos Maciel and R. Ballini, On the predictability of high and low prices: The case of bitcoin, Brazilian Review of Finance, 17(3), pp. 66–84, 2019.

[23] R. F. Engle and C. W. Granger, Co-integration and error correction: representation, estimation, and testing, Econometrica: journal of the Econometric Society, pp. 251–276, 1987.

[24] D. Fantazzini and N. Kolodin, Does the hashrate affect the bitcoin price?, Journal of Risk and Financial Management, 13(11), p. 263, 2020.

[25] D. Garcia, C. J. Tessone, P. Mavrodiev, and N. Perony, The digital traces of bubbles: feedback cycles between socio-economic signals in the bitcoin economy, Journal of the Royal Society Interface, 11(99), p. 20140623, 2014.

[26] E. Gemici and M. Polat, Relationship between price and volume in the bitcoin market, The Journal of Risk Finance, 2019.

[27] I. Georgoula, D. Pournarakis, C. Bilanakos, D. Sotiropoulos, and G. M. Giaglis, Using time-series and sentiment analysis to detect the determinants of bitcoin prices, Available at SSRN 2607167, 2015.

[28] J.-C. Gerlach, G. Demos, and D. Sornette, Dissection of bitcoin's multiscale bubble history from january 2012 to february 2018, Royal Society open science, 6(7), p. 180643, 2019.

[29] B. Gianmarco, *A new model for bitcoin mining costs: an econometric analysis of the bitcoin price and cost dynamics*, Master's thesis, Politecnico di Milano, Politecnico di Milano Piazza Leonardo da Vinci, 32 20133 Milano, 2022.

[30] C. W. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica: journal of the Econometric Society, pp. 424–438, 1969.

[31] A. Greaves and B. Au, Using the bitcoin transaction graph to predict the price of bitcoin, No data, 8, pp. 416–443, 2015.

[32] F. E. Harrell et al., *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*, volume 608, Springer, 2001.

[33] A. S. Hayes, Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin, Telematics and informatics, 34(7), pp. 1308–1321, 2017.

[34] C. C. Holt, Forecasting seasonals and trends by exponentially weighted moving averages, International journal of forecasting, 20(1), pp. 5–10, 2004.

[35] R. Hyndman, A. B. Koehler, J. K. Ord, and R. D. Snyder, *Forecasting with exponential smoothing: the state space approach*, Springer Science & Business Media, 2008.

[36] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*, OTexts, 2018.

[37] A. Ibrahim, R. Kashef, and L. Corrigan, Predicting market movement direction for bitcoin: A comparison of time series modeling methods, Computers & Electrical Engineering, 89, p. 106905, 2021.

[38] S. Ji, J. Kim, and H. Im, A comparative study of bitcoin price prediction using deep learning, Mathematics, 7(10), p. 898, 2019.

[39] S. Johansen, Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models, Econometrica: journal of the Econometric Society, pp. 1551–1580, 1991.

[40] S. H. Kang, R. P. McIver, and J. A. Hernandez, Co-movements between bitcoin and gold: A wavelet coherence analysis, Physica A: Statistical Mechanics and its Applications, 536, p. 120888, 2019.

[41] B. Kapar and J. Olmo, An analysis of price discovery between bitcoin futures and spot markets, Economics Letters, 174, pp. 62–64, 2019.

[42] B. Kapar and J. Olmo, Analysis of bitcoin prices using market and sentiment variables, The World Economy, 44(1), pp. 45–63, 2021.

[43] A. M. Khedr, I. Arif, M. El-Bannany, S. M. Alhashmi, and M. Sreedharan, Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey, Intelligent Systems in Accounting, Finance and Management, 28(1), pp. 3–34, 2021.

[44] F. Kjærland, A. Khazal, E. A. Krogstad, F. B. Nordstrøm, and A. Oust, An analysis of bitcoin's price dynamics, Journal of Risk and Financial Management, 11(4), p. 63, 2018.

[45] Y. Lee and J. H. Rhee, A vecm analysis of bitcoin price using time-varying cointegration approach, Journal of Derivatives and Quantitative Studies, 30(1), 2022.

[46] H. Lütkepohl, *New introduction to multiple time series analysis*, Springer Science & Business Media, 2005.

[47] Š. Lyócsa, P. Molnár, T. Plíhal, and M. Širaňová, Impact of macroeconomic news, regulation and hacking exchange markets on the volatility of bitcoin, Journal of Economic Dynamics and Control, 119, p. 103980, 2020.

[48] K. B. Management, Grenoble Ecole de, Block Chain 2.0: The Renaissance of Money, Wired, January 2015, ISSN 1059-1028, section: Innovation Insights.

[49] A. Meynkhard, Fair market value of bitcoin: Halving effect, Investment Management and Financial Innovations, 16(4), pp. 72–85, 2019.

[50] M. Mudassir, S. Bennbaia, D. Unal, and M. Hammoudeh, Time-series forecasting of bitcoin prices using high-dimensional features: a machine learning approach, Neural computing and applications, pp. 1–15, 2020.

[51] S. Nakamoto, Bitcoin: A peer-to-peer electronic cash system, Decentralized Business Review, 2008.

[52] D. Philippas, H. Rjiba, K. Guesmi, and S. Goutte, Media attention and bitcoin prices, Finance Research Letters, 30, pp. 37–43, 2019.

[53] P. C. Phillips and B. E. Hansen, Statistical inference in instrumental variables regression with i (1) processes, The Review of Economic Studies, 57(1), pp. 99–125, 1990.

[54] G. Pieters and S. Vivanco, Financial regulations and price inconsistencies across bitcoin markets, Information Economics and Policy, 39, pp. 1–14, 2017.

[55] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022.

[56] S. Raju and A. M. Tarif, Real-time prediction of bitcoin price using machine learning techniques and public sentiment analysis, arXiv preprint arXiv:2006.14473, 2020.

[57] C. A. Sims, Comparison of interwar and postwar business cycles: Monetarism reconsidered, 1980.

[58] J. H. Stock and M. W. Watson, Testing for common trends, Journal of the American statistical Association, 83(404), pp. 1097–1107, 1988.

[59] S. J. Taylor and B. Letham, Forecasting at scale, The American Statistician, 72(1), pp. 37–45, 2018.

[60] J. Truby, Decarbonizing bitcoin: Law and policy choices for reducing the energy consumption of blockchain technologies and digital currencies, Energy research & social science, 44, pp. 399–410, 2018.

[61] N. Uras, L. Marchesi, M. Marchesi, and R. Tonelli, Forecasting bitcoin closing price series using linear regression and neural networks models, PeerJ Computer Science, 6, p. e279, 2020.

[62] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*, CreateSpace, Scotts Valley, CA, 2009, ISBN 1441412697.

[63] J. Wang, Y. Xue, and M. Liu, An analysis of bitcoin price based on vec model, in *2016 international conference on economics and management innovations*, pp. 180–186, Atlantis Press, 2016.

[64] P. R. Winters, Forecasting sales by exponentially weighted moving averages, Management science, 6(3), pp. 324–342, 1960.

[65] I. Yenidoğan, A. Çayir, O. Kozan, T. Dağ, and Ç. Arslan, Bitcoin forecasting using arima and prophet, in *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, pp. 621–624, IEEE, 2018.

[66] G. U. Yule, Why do we sometimes get nonsense-correlations between time-series?–a study in sampling and the nature of time-series, Journal of the royal statistical society, 89(1), pp. 1–63, 1926.

[67] G. U. Yule, Vii. on a method of investigating periodicities disturbed series, with special reference to wolfer's sunspot numbers, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 226(636-646), pp. 267–298, 1927.

[68] Y. Zhu, D. Dickinson, and J. Li, Analysis on the influence factors of bitcoin's price based on vec model, Financial Innovation, 3(1), pp. 1–13, 2017.

APPENDIX A

VISUALIZATION OF THE VARIABLES

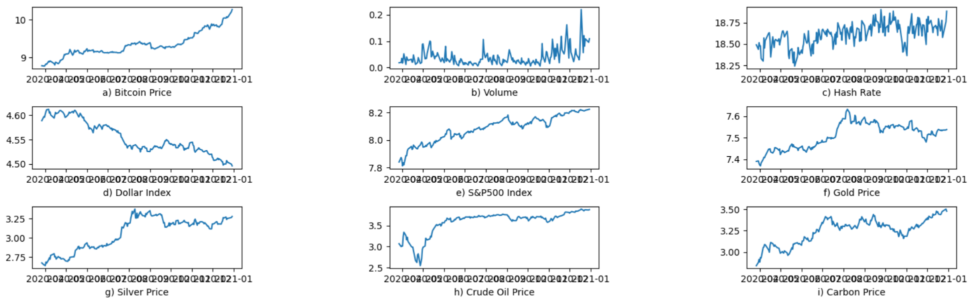


Figure A.1: Period 2: Visualization of transformed variables

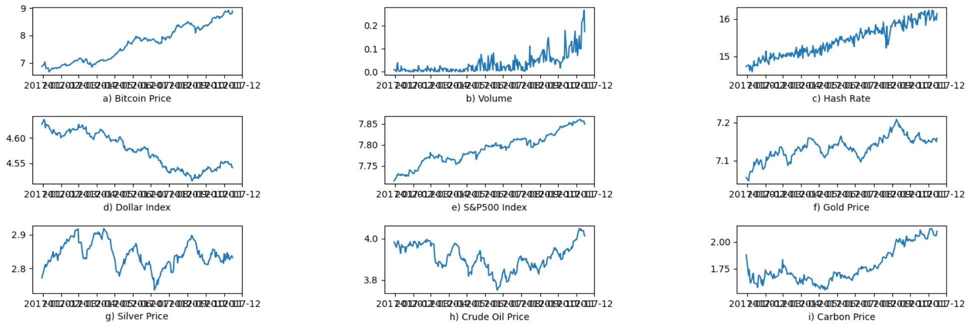


Figure A.2: Period 3: Visualization of transformed variables

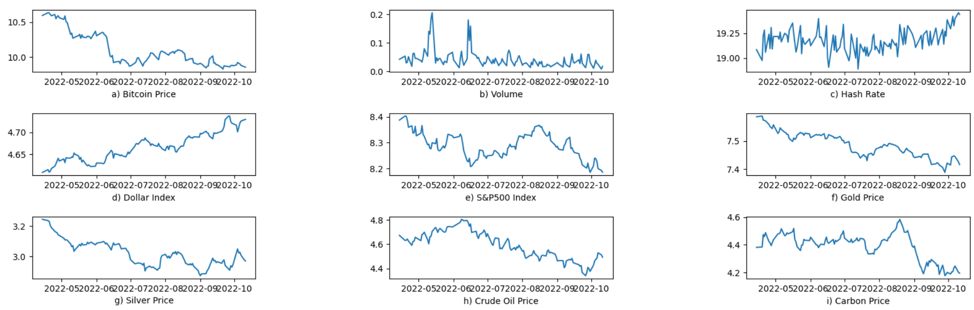


Figure A.3: Period 4: Visualization of transformed variables

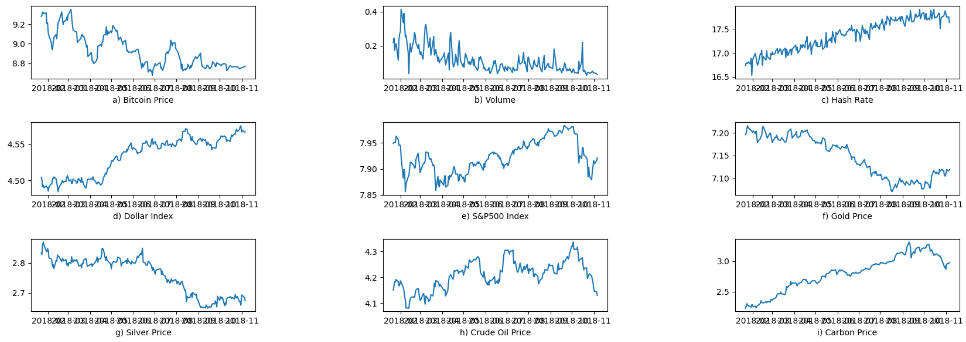


Figure A.4: Period 5: Visualization of transformed variables

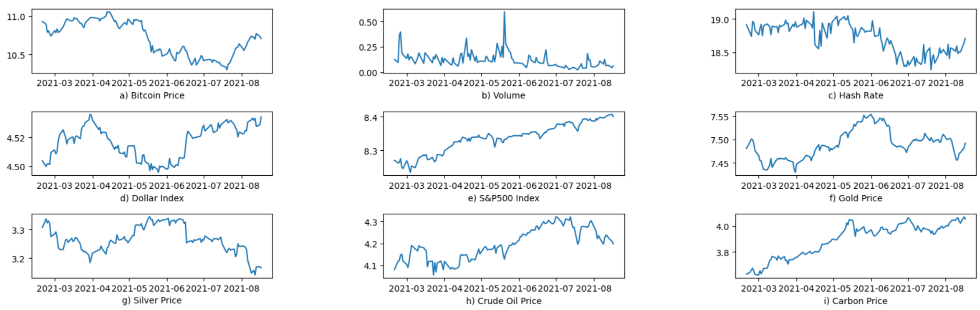


Figure A.5: Period 6: Visualization of transformed variables

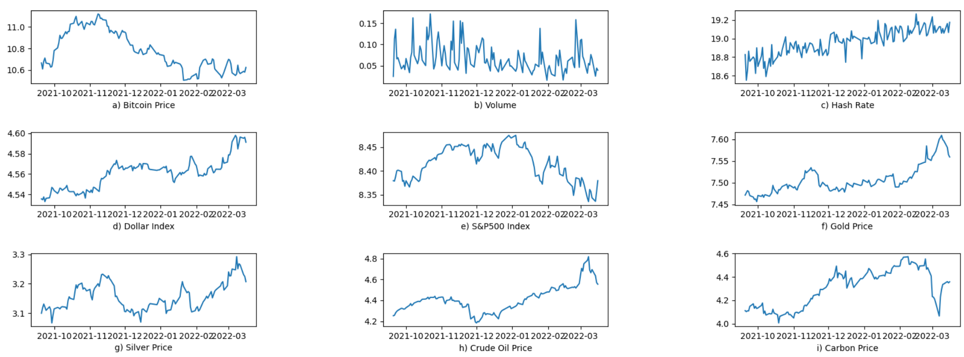


Figure A.6: Period 7: Visualization of transformed variables

APPENDIX B

SMOOTHING MODELS

Table B.1: Period 1: Model coefficients of ES models

	SES		Holt's Linear				Holt's Exponential			
	α	**initial level**	α	β	**initial level**	**initial trend**	α	β	**initial level**	**initial trend**
BTCPrice	1,000	5,922	1,000	0,000	5,919	0.003	1,000	0,000	5,920	1,000
Hashrate	0,286	14,003	0,250	0,006	13,980	0,004	0,249	0,007	13,981	1,000
Volume	0,356	0,004	0,356	0,000	0,004	0,000	0,374	0,000	0,004	0,983
Gold	1,000	7,081	1,000	0,000	7,080	0,000	1,000	0,000	7,080	1,000
Silver	0,926	2,737	0,926	0,000	2,737	0,000	0,926	0,000	2,737	1,000
Crudeoil	1,000	3,330	1,000	0,000	3,329	0,001	1,000	0,000	3,329	1,000
DXY	0,995	4,565	0,995	0,000	4,565	0,000	0,995	0,000	4,566	0,999
SP500	0,961	7,524	0,960	0,000	7,523	0,001	0,959	0,000	7,523	1,000
Carbon	0,960	1,598	0,957	0,000	1,596	0,002	0,957	0,000	1,597	1,000

APPENDIX C

ARIMA MODELS

C.1 Period 1

Table C.1: Period 1: Model output of $ARIMA(0, 1, 1)$ for Hashrate

	Estimate	Std. Error	z	p-value
intercept	0.0035	0.001	4.997	0.000
ma1	-0.7412	0.016	-47.347	0.000
sigma2	0.0095	0.000	29.250	0.000

Log-likelihood	1227.557	AIC	-2549.155
BIC	-2533.413	HQIC	-2543.271

	z	p-value
Ljung-Box	0.04	0.84
Heteroskedasticity	0.94	0.52
Jarque-Bera (JB)	49.40	0.00

Table C.2: Period 1: Model output of $ARIMA(4, 1, 5)$ for Volume

	Estimate	Std. Error	z	p-value
ar1	-0.2794	0.212	-1.318	0.187
ar2	-0.5277	0.066	-8.011	0.000
ar3	-0.1875	0.181	-1.036	0.300
ar4	0.3611	0.069	5.248	0.000
ma1	-0.2743	0.212	-1.294	0.196
ma2	0.2546	0.152	1.678	0.093
ma3	-0.1616	0.165	-0.977	0.328
ma4	-0.6508	0.152	-4.270	0.000
ma5	0.1420	0.045	3.161	0.002
sigma2	0.0023	0.000	90.182	0.000

Log-likelihood	2271.569	AIC	-4523.137
BIC	-4470.667	HQIC	-4503.525

	z	p-value
Ljung-Box	0.08	0.78
Heteroskedasticity	0.85	0.08
Jarque-Bera (JB)	13175.70	0.00

Table C.3: Period 1: Model output of $ARIMA(0, 1, 0)$ for Gold

	Estimate	Std. Error	z	p-value
intercept	0.0003	0.000	1.291	0.197
sigma2	0.0001	0.000	38.610	0.000

Log-likelihood	4717.173	AIC	-9430.345
BIC	-9419.851	HQIC	-9426.423

	z	p-value
Ljung-Box	1.07	0.30
Heteroskedasticity	1.32	0.00
Jarque-Bera (JB)	300.29	0.00

Table C.4: Period 1: Model output of $ARIMA(0, 1, 2)$ for Silver

	Estimate	Std. Error	z	p-value
ma1	-0.0757	0.017	-4.414	0.000
ma2	0.0487	0.018	2.638	0.008
sigma2	0.0003	0.000	57.076	0.000

Log-likelihood	3685.281	AIC	-7364.563
BIC	-7348.821	HQIC	-7358.679

	z	p-value
Ljung-Box	0.00	0.95
Heteroskedasticity	2.78	0.00
Jarque-Bera (JB)	3542.85	0.00

Table C.5: Period 1: Model output of $ARIMA(5, 1, 2)$ for Crudeoil

	Estimate	Std. Error	z	p-value
ar1	0.0876	0.058	1.513	0.130
ar2	-0.7591	0.044	-17.200	0.000
ar3	0.1735	0.021	8.209	0.000
ar4	0.0729	0.023	3.216	0.001
ar5	-0.0132	0.017	-0.756	0.450
ma1	0.0497	0.056	0.896	0.370
ma2	0.7378	0.040	18.416	0.000
sigma2	0.0008	0.000	68.371	0.000

Log-likelihood	3001.870	AIC	-5987.739
BIC	-5945.763	HQIC	-5972.050

	z	p-value
Ljung-Box	0.00	0.95
Heteroskedasticity	2.85	0.00
Jarque-Bera (JB)	2155.57	0.00

Table C.6: Period 1: Model output of $ARIMA(0,1,0)$ for DXY

	Estimate	Std. Error	z	p-value
sigma2	0.0001	0.000	34.615	0.000

Log-likelihood	5842.596	AIC	-11683.192
BIC	-11677.945	HQIC	-11681.231

	z	p-value
Ljung-Box	0.01	0.94
Heteroskedasticity	0.74	0.00
Jarque-Bera (JB)	119.55	0.00

Table C.7: Period 1: Model output of $ARIMA(3,1,0)$ for SP500

	Estimate	Std. Error	z	p-value
intercept	0.0007	0.000	2.283	0.022
ar1	-0.0353	0.015	-2.291	0.022
ar2	0.0954	0.015	6.411	0.000
ar3	-0.0695	0.015	-4.756	0.000
sigma2	0.0001	0.000	53.693	0.000

Log-likelihood	4519.375	AIC	-9028.751
BIC	-9002.515	HQIC	-9018.945

	z	p-value
Ljung-Box	0.01	0.92
Heteroskedasticity	4.37	0.00
Jarque-Bera (JB)	4087.69	0.00

Table C.8: Period 1: Model output of $ARIMA(1,1,1)$ for Carbon

	Estimate	Std. Error	z	p-value
intercept	0.0030	0.001	2.155	0.031
ar1	-0.6964	0.211	-3.297	0.001
ma1	0.6539	0.221	2.961	0.003
sigma2	0.0009	0.000	48.071	0.000

Log-likelihood	2965.103	AIC	-5922.206
BIC	-5901.218	HQIC	-5914.361

	z	p-value
Ljung-Box	0.00	0.98
Heteroskedasticity	0.77	0.01
Jarque-Bera (JB)	1452.47	0.00

C.2 Period 2

Table C.9: Period 2: Model output of $ARIMA(1,1,0)$ for BTC

	Estimate	Std. Error	z	p-value
intercept	0.0093	0.003	3.252	0.001
ar1	-0.1636	0.061	-2.685	0.007
sigma2	0.0014	0.000	15.388	0.000

Log-likelihood	346.831	AIC	-687.663
BIC	-677.986	HQIC	-683.741

	z	p-value
Ljung-Box	0.00	0.95
Heteroskedasticity	0.98	0.93
Jarque-Bera (JB)	88.95	0.00

Table C.10: Period 2: Model output of $ARIMA(0,1,1)$ for Hashrate

	Estimate	Std. Error	z	p-value
ma1	-0.6608	0.053	-12.377	0.000
sigma2	0.0092	0.001	9.139	0.000

Log-likelihood	171.952	AIC	-339.904
BIC	-333.453	HQIC	-337.290

	z	p-value
Ljung-Box	0.06	0.81
Heteroskedasticity	1.28	0.33
Jarque-Bera (JB)	1.88	0.39

Table C.11: Period 2: Model output of $ARIMA(1,1,1)$ for Volume

	Estimate	Std. Error	z	p-value
ar1	0.1915	0.062	3.113	0.002
ma1	-0.8540	0.045	-18.797	0.000
sigma2	0.0006	0.000	16.630	0.000

Log-likelihood	422.337	AIC	-838.675
BIC	-828.998	HQIC	-834.753

	z	p-value
Ljung-Box	0.02	0.90
Heteroskedasticity	3.12	0.00
Jarque-Bera (JB)	360.93	0.00

Table C.12: Period 2: Model output of $ARIMA(0, 1, 0)$ for Gold

	Estimate	Std. Error	z	p-value
sigma2	0.0001	0.000	13.109	0.000

Log-likelihood	577.985	AIC	-1153.969
BIC	-1150.744	HQIC	-1152.662

	z	p-value
Ljung-Box	0.19	0.66
Heteroskedasticity	1.03	0.90
Jarque-Bera (JB)	34.58	0.00

Table C.13: Period 2: Model output of $ARIMA(0, 1, 2)$ for Silver

	Estimate	Std. Error	z	p-value
ma1	-0.1594	0.066	-2.409	0.016
ma2	0.1654	0.060	2.751	0.006
sigma2	0.0008	0.000	13.168	0.000

Log-likelihood	401.208	AIC	-796.415
BIC	-786.738	HQIC	-792.494

	z	p-value
Ljung-Box	0.01	0.91
Heteroskedasticity	1.22	0.43
Jarque-Bera (JB)	38.70	0.00

Table C.14: Period 2: Model output of $ARIMA(0, 1, 4)$ for Crudeoil

	Estimate	Std. Error	z	p-value
ma1	0.2032	0.053	3.811	0.000
ma2	-0.2033	0.045	-4.508	0.000
ma3	0.1295	0.055	2.362	0.018
ma4	0.3637	0.057	6.409	0.000
sigma2	0.0023	0.000	15.706	0.000

Log-likelihood	301.934	AIC	-593.867
BIC	-577.739	HQIC	-587.331

	z	p-value
Ljung-Box	0.04	0.84
Heteroskedasticity	0.13	0.00
Jarque-Bera (JB)	327.47	0.00

Table C.15: Period 2: Model output of $ARIMA(0, 1, 0)$ for DXY

	Estimate	Std. Error	z	p-value
intercept	-0.0005	0.000	-1.802	0.072
sigma2	0.0001	0.000	8.661	0.000

Log-likelihood	775.109	AIC	-1546.218
BIC	-1539.766	HQIC	-1543.603

	z	p-value
Ljung-Box	0.10	0.75
Heteroskedasticity	0.57	0.03
Jarque-Bera (JB)	1.38	0.50

Table C.16: Period 2: Model output of $ARIMA(3, 0, 2)$ for SP500

	Estimate	Std. Error	z	p-value
intercept	0.1030	0.008	12.640	0.000
ar1	0.1136	0.143	0.795	0.427
ar2	0.5342	0.056	9.588	0.000
ar3	0.3394	0.153	2.220	0.026
ma1	0.7935	0.111	7.148	0.000
ma2	0.4249	0.122	3.491	0.000
sigma2	0.0002	0.000	11.191	0.000

Log-likelihood	529.879	AIC	-1045.758
BIC	-1023.141	HQIC	-1036.594

	z	p-value
Ljung-Box	0.33	0.57
Heteroskedasticity	0.27	0.00
Jarque-Bera (JB)	104.11	0.00

Table C.17: Period 2: Model output of $ARIMA(1, 1, 1)$ for Carbon

	Estimate	Std. Error	z	p-value
intercept	0.0061	0.004	1.723	0.085
ar1	-0.7515	0.169	-4.441	0.000
ma1	0.5905	0.210	2.818	0.005
sigma2	0.0009	0.000	10.991	0.000

Log-likelihood	391.942	AIC	-775.884
BIC	-762.981	HQIC	-770.655

	z	p-value
Ljung-Box	0.01	0.92
Heteroskedasticity	0.49	0.01
Jarque-Bera (JB)	8.92	0.01