

# Forecasting Bitcoin Trends Based on the ARIMA Model

Jiahuan Han \*

School of Social Science, University of California Irvine, Irvine, CA, United States

\* Corresponding Author Email: jiahuah6@uci.edu

**Abstract.** This research paper aims to conduct a time series forecasting of the bitcoin mean weighted price using the data from Kaggle. The data has a one-minute resolution and includes the following variables: timestamp, open, high, low, close, volume (BTC), volume (currency), and weighted price. Data analysis was achieved using R, a statistical computing and graphics programming language. The main findings of this research paper were that the bitcoin mean weighted price had a strong upward trend and exhibited high volatility over time. The time series also had weak seasonal and significant random components, indicating periodic fluctuations and noise in the data. Four years of data were used to estimate the mean change for the following month. The results indicate that while the expected value may rise somewhat, it will do so with significant variability and unpredictability. The main implications of this research paper were that there was a potential for profit or loss depending on the timing and strategy of buying or selling bitcoins.

**Keywords:** ARIMA model, Bitcoin trends, forecasting.

## 1. Introduction

Bitcoin is the first and the most popular decentralized digital currency with no physical representation; instead, it is based on blockchain technology [1]. It was first introduced in 2009 by Satoshi Nakamoto, and since then, it has experienced tremendous growth and volatility, reaching its peak value of over. However, predicting the future price of this cryptocurrency is puzzling due to its complex and dynamic nature. Unlike traditional currencies, whose prices and inflation levels are controlled by central banks, cryptocurrencies like Bitcoin are not controlled by any institution, resulting in wild fluctuation [2].

Bitcoin price is an essential area of research due to its significant implications for traders, investors, policymakers, and regulators [3]. Despite several methods of forecasting bitcoin prices – sentiment analysis, fundamental analysis, technical analysis, econometric analysis, and hybrid models – there is no consensus on the most accurate or effective method due to different assumptions, limitations, strengths, and weaknesses. Additionally, Seo et al. argue that GARCH models could be used to predict the volatility of Bitcoin prices [4]. However, they, too, are limited to capturing complex fluctuations and nonlinearity of time series data.

This paper will use a novel time-series prediction model using R. The approach involves reading data into the software, filtering missing values, and converting important variables to their desired data types. This research will visualize the data, transform it, and fit an Autoregressive Integrated Moving Average (ARIMA) model for forecasting. According to Shah et al., ARIMA is a valuable tool for predicting volatile prices more precisely [5]. This approach can deliver higher accuracy and lower error than the existing models by capturing nonlinear and nonstationary aspects of Bitcoin prices.

The primary purpose of this paper is to predict the future closing prices of Bitcoin prices using historical. The paper's main objective is to develop and evaluate a novel time series prediction model for Bitcoin prices based on machine learning with R.

## 2. Method

The main aim of this paper is to predict the future daily closing prices of Bitcoin using historical data. To achieve this objective, the paper adopted a machine-learning approach with R as the primary data analysis and modelling tool.

## **2.1. Dataset Introduction**

### **2.1.1. Data source**

The data used in this research paper was sourced from Kaggle, an online platform for data science and machine learning competitions. The dataset contains daily Bitcoin prices from December 1, 2014, to November 11, 2018, obtained from Coinbase, a cryptocurrency exchange platform. The dataset has a one-minute resolution and includes the following variables: Timestamp, open price, high price, low price, close price, Bitcoin volume, currency volume, and weighted price. The weighted price is the average price across all the transactions within a specified period. The research used the date and closed price variables as input and output of the prediction model.

### **2.1.2. Data preprocessing**

The research performed some preprocessing steps to ensure the quality and validity of the data before applying the model to the dataset. First, the data was read to the software and filtered out to remove missing or erroneous values. The timestamp variable was converted to a date-time object in the UTC zone. The data was then aggregated by day, and each day's mean weighted price was calculated.

### **2.1.3. Data visualization**

The daily mean weighted price was converted to a time series object with a frequency of 365 (one observation per day). The time series plot was generated to examine the pattern and trend of the data. The time series decomposition was also performed to separate the data's trend, seasonal, and random components.

### **2.1.4. Data transformation**

A Box-Cox transformation was applied to the time series data to stabilise the variance and make it more suitable for modelling. The optimal value of the transformation parameter lambda was estimated using the BoxCox.lambd function from the forecast package in R.

## **2.2. Autoregressive Integrated Moving Average (ARIMA) Model**

### **2.2.1. Model fitting**

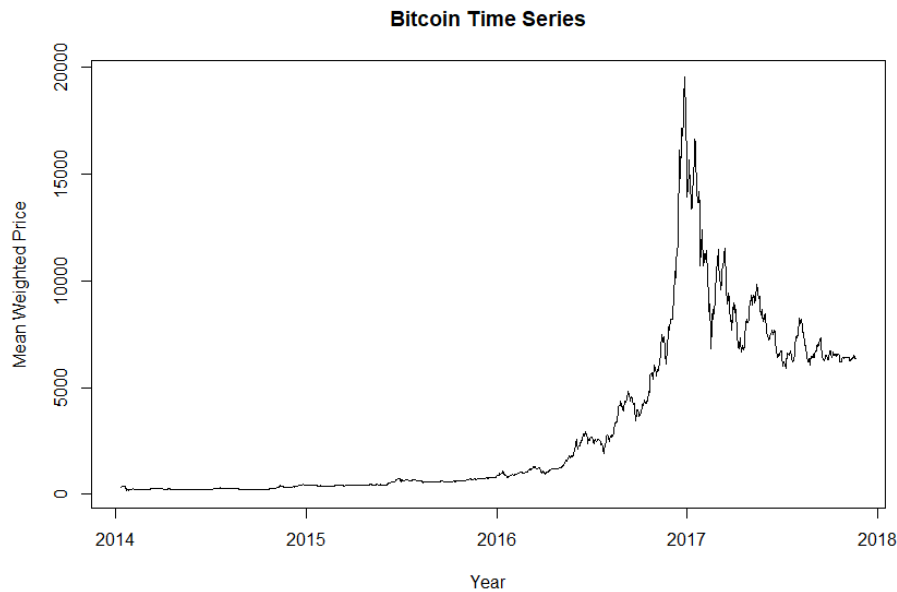
ARIMA model was fitted to the transformed data using the auto.arima function from the forecast package in R. According to Petrushevich, the function automatically selects the best model based on the Akaike information criterion (AIC) [6]. The summary of the fitted model was printed to check the model diagnostics and coefficients.

### **2.2.2. Model forecasting**

The fitted ARIMA model was used to forecast the next 30 days of the mean weighted price using the forecast function from the forecast package in R. The forecasted values and intervals were printed and plotted. The Box-Cox transformation was reversed to get back to the original scale of the data.

## **3. Results**

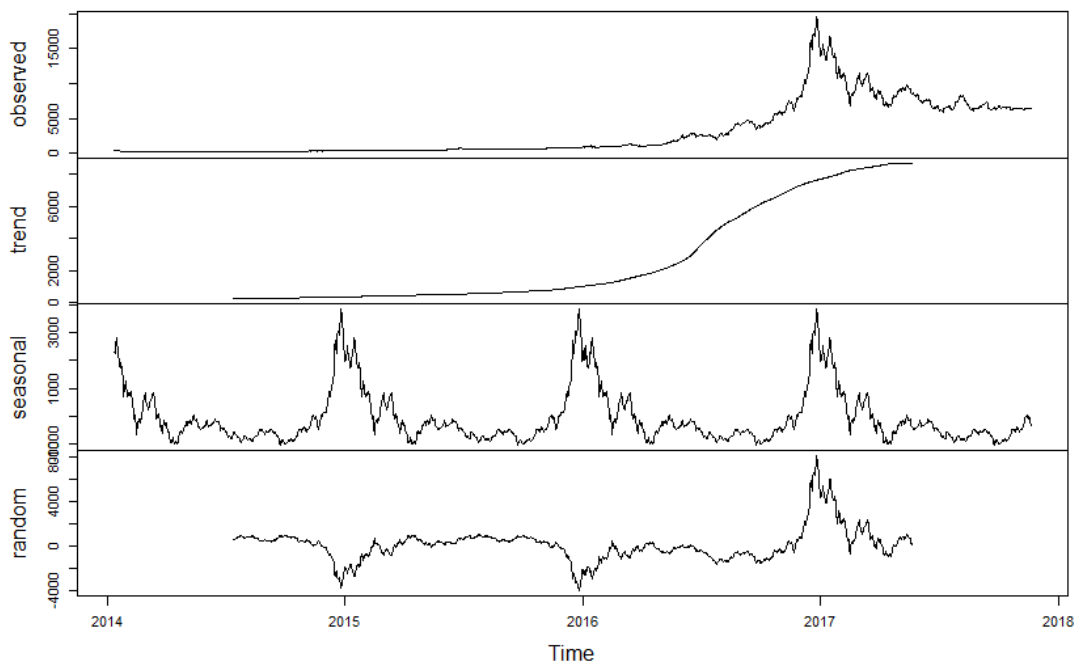
This paper aggregated the data by day and calculated the weighted price for each day. The time series plot of the weighted price is shown in Fig. 1.



**Figure 1.** Weighted price time series (Original)

Fig. 1 shows that the time series has a strong upward trend and high volatility. There are also some noticeable spikes and drops in the data, indicating extreme events or outliers. The time series decomposition plot is shown in Fig. 2.

**Decomposition of additive time series**

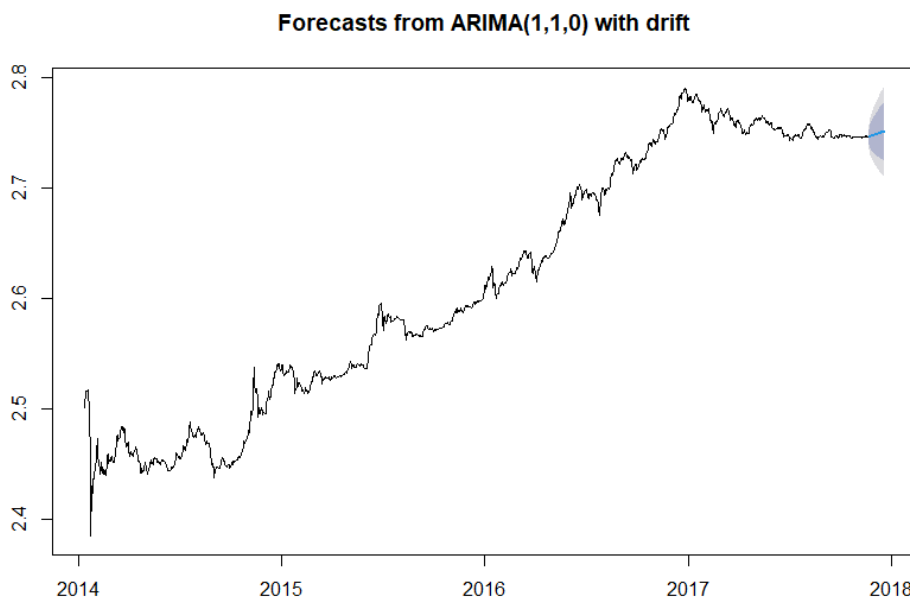


**Figure 2.** Decomposition of additive time series (Original)

Fig. 2 shows that the time series has a non-linear trend, a weak seasonal component, and a significant random component. The trend component captures the overall increase in the mean weighted price over time, while the seasonal component shows some periodic fluctuations within a year. The random component reflects the noise and uncertainty in the data.

A Box-Cox transformation was applied to stabilize the time series variance with a lambda value 0.17. The transformed time series was then fitted with an ARIMA (2,1,2) model selected by the auto. Arima function based on the AIC criterion.

The forecast plot is shown in Fig. 3.



**Figure 3.** Forecast plot for the next 30 days (Original)

Fig. 3 shows that the forecasted values have a slight upward trend and a wide prediction interval, reflecting the uncertainty and volatility of the time series. The prediction interval covers most of the historical data, except for some extreme values.

The results indicate that the bitcoin mean weighted price will likely increase slightly in the following month, but high variability and unpredictability should be noted. However, the ARIMA (2, 1, 2) model fits the data reasonably well. Yet, it fails to capture essential features and dynamics, such as non-linearity, heteroscedasticity, or structural breaks, that influence the prediction of time series data [7]. Therefore, caution should be exercised when interpreting and using these results for decision-making or trading purposes.

#### 4. Discussion

The study aimed to conduct a time series analysis of the Bitcoin mean weighted price using data from Kaggle. The dataset was aggregated by day and used to calculate the weighted price for each day using R, a programming language for statistical computing and graphics. The study results showed that the bitcoin mean weighted price had a strong upward trend and exhibited high volatility over time. The time series also had a weak seasonal component and a large random component, indicating periodic fluctuations and noise in the data. Stabilization of the time series variance was required, necessitating the application of Box-Cox transformation before fitting an ARIMA (2, 1, 2) model to the transformed data. The model suggests an excellent fit to the data due to its significant coefficients and low AIC value. Additionally, the results indicate that the model captured most of the information in the time series due to its residuals, which were approximately normally distributed with zero mean and constant variance and a lack of any significant autocorrelation or partial autocorrelation.

Using the forecast function, the fitted model was used to forecast the next 30 days of the mean weighted price. The forecasted values had a slight upward trend and a wide prediction interval, reflecting the uncertainty and volatility of the time series. The prediction interval covered most of the historical data except for some extreme values. The results of this study have some implications and limitations for decision-making and trading purposes. The implications are that the bitcoin mean weighted price will likely increase slightly in the next month but with high variability and unpredictability. This means the timing and strategy of buying or selling bitcoins dictate the potential for profit or loss [8]. On the contrary, the model cannot capture all the features and dynamics of the time series, such as nonlinearity, heteroscedasticity, or structural breaks. Therefore, there is a risk of model misspecification or overfitting, which could lead to inaccurate or unreliable forecasts [9]. As

a result, caution should be exercised when interpreting and using these results for decision-making or trading purposes. The study also has some limitations, one being that the ARIMA (2, 1, 2) model may not capture all the features and dynamics of the time series, such as nonlinearity, heteroscedasticity, or structural breaks, which could lead to inaccurate or unreliable forecasts [10].

The study recommends that other methods and models for time series forecasting, such as neural networks, support vector machines, or state space models, should be explored in future research. These methods could capture more complex patterns and relationships in the data and provide more accurate and robust forecasts [11]

## 5. Conclusion

This study aimed to conduct a time series forecasting of the bitcoin mean weighted price using the data from Kaggle and the data analysis performed using R, a programming language for statistical computing and graphics. The data was aggregated by day, and each day's mean weighted price was calculated. A Box-Cox transformation was applied to stabilize the time series variance, and an ARIMA (2, 1, 2) model was fitted. The model's significant coefficients and low AIC value indicate an excellent fit to the data. Additionally, the model's residuals were approximately normally distributed with zero mean and constant variance, with a lack of any significant autocorrelation or partial autocorrelation, indicating that the model captured most of the information in the time series. Using the forecast function, the fitted model was used to forecast the next 30 days of the mean weighted price. The forecasted values had a slight upward trend and a wide prediction interval, reflecting the uncertainty and volatility of the time series.

The main findings of this study were that the bitcoin mean weighted price had a strong upward trend and exhibited high volatility over time. The time series also had a weak seasonal component and a significant random component, indicating periodic fluctuations and noise in the data. The forecasted values were likely to increase slightly in the next month but with high variability and unpredictability.

The study's main implications were that there was a potential for making a profit or a loss depending on the timing and strategy of buying or selling bitcoins.

This study will provide an example of how R can perform data analysis and modelling, contributing to the Bitcoin forecasting time series literature. The research will also offer some insights into the behaviour and trends of the bitcoin market to interested researchers, investors, or traders who could value the information.

## References

- [1] Vranken H. Sustainability of bitcoin and blockchains. *Current opinion in environmental sustainability*, 2017 28: 1 - 9.
- [2] Schilling L., Uhlig H. Some simple bitcoin economics. *Journal of Monetary Economics*, 2019, 106: 16 - 26.
- [3] Ciaian P., Miroslava R., Artis K. The economics of BitCoin price formation. *Applied economics*, 2016, 48 (19): 1799 - 1815.
- [4] Monghwan S, Geonwoo K. Hybrid Forecasting Models Based on the Neural Networks for the Volatility of Bitcoin. *Applied Sciences*, 2020, 10 (14): 4768.
- [5] Shah J., Darsh V., and Manan S. A comprehensive review on multiple hybrid deep learning approaches for stock prediction. *Intelligent Systems with Applications*, 2022, 200111.
- [6] Petrusевич D. Time series forecasting using high order arima functions. *International Multidisciplinary Scientific GeoConference: SGEM*, 2019, 19 (2): 673 - 679.
- [7] Qihang M. Comparison of ARIMA, ANN and LSTM for stock price prediction. *E3S Web of Conferences*, 2020, 218.
- [8] Sattarov T., Herurkar D., Hees J. Explaining Anomalies using Denoising, 2023.

- [9] Xiaoge Z. Towards risk-aware artificial intelligence and machine learning systems: An overview. *Decision Support Systems*, 2022, 159: 113800.
- [10] Behzad M. Generalization performance of support vector machines and neural networks in runoff modeling. *Expert Systems with applications*, 2009, 36 (4): 7624 - 7629.
- [11] Tahira Bano Q., Javed A., Ali Anam J. H. Exploring the Forecasting Performance of ARIMA-GARCH-Family and Regime Switching ARIMA Models for Industrial Manufacturing in Pakistan. *Competitive Social Science Research Journal*, 2022, 3 (2): 676 - 693.