Forecasting Bitcoin Prices using Time Series Analysis

Alimzhan Gumran, Anel Salmenova, Akezhan Shakenov
Piotr Sebastian Skrzypacz
Nazarbayev University

December 1, 2024

Abstract

The study analyzes the Bitcoin Historical dataset and evaluates it. Bitcoin is one of the leading cryptocurrencies, and forecasting its future price is of special interest for cryptocurrency investors. In the scope of this research project the prices of bitcoin was considered to be a set of random variable and, thus, treated as stochastic process. This allowed to use time series analysis methods and build effective models that can predict bitcoin prices based on prior values. Three different models were built based on the prior prices, namely AR (1), AR (2) and ARMA (1,1) that proved to be effective because of low computational costs – models do not include more than 2 estimators. This paper reveals the comparison of the models in terms of the most accurate result.

Keywords: Bitcoin, Linear Models, Model Comparison, Prediction Accuracy, AR(1), AR(2), ARMA(1,1).

1 Introduction

Cryptocurrency is, in its essence, a decentralized digital exchange tool. Its exchange is based on blockchain technology that keeps every exchange in the specially designed repository [1]. The primary goal of this report is to identify the most accurate model for bitcoin price forecasting. This problem is of interest for those interested in investing in the cryptocurrency. Cryptocurrency as any stock has continuously changing price which can raise or decay. In either case accurate forecast will help investor to adjust their strategy. When the price is expected to go down, shorting strategies are used; and long strategy is used when price is expected to raise. That makes our project especially relevant today as according to Yahoo Finance the closing price of Bitcoin was around 38600 in December 2023 and approximately 95900 in December 2024. The fact that almost 3 times growth in one year is not observed in any other currency shows obvious interest in being able to track what the price is going to be. In the investment field it is commonly known that price incorporates every factor that affects the price. As price is a comprehensive estimator for future values, using autoregressive and autoregressive moving-average models is suitable in this topic case.

2 Research Framework

2.1 Problem Overview

The project aims to test different models and identify the most optimal one that has the results closest to the actual data. The dataset contains characteristics, including the time of the stamp, highest, lowest, closest values, volumes in Bitcoin and US Dollars. Using these data variables, we needed to define in our model what we consider to be an ultimate price and then we build our model relatively to it. In order to build our model we made appropriate estimations for AR(1), AR(2) and ARMA(1,1) models. We chose those for their simplicity and computational efficiency. We visualized our data and did necessary computations using Python programming language.

2.2 Methodology

We used publicly available Kaggle Bitcoin historical dataset that shows minute, hourly, and daily interval data. We divided it into train and test datasets to evaluate the performance of our predictive models and prevent overfitting. The training dataset is used to fit and optimize the parameters of the model, whereas the testing one assesses how well the models behave (generalize) behind the unseen data.

2.3 Modeling

Our choice fell on the AR (1), AR (2) and ARMA (1,1) models as they are a relevant option for time series data analysis. AR(1) and AR(2) models provide a relatively simple and straightforward approach to capturing the dependency of Bitcoin prices, where the current value is influenced by one or two lagged values. And the ARMA(1,1) model extends this by incorporating both autoregressive and moving average components, allowing us to see more complex patterns [4]. Moreover, they minimize computational cost and produce less error. These features meet the specific needs for our project and make the models ideal for exploring the behavior of Bitcoin price fluctuations [6].

For our project, a price per a minute t is the average of lowest and highest at that minute t. By this, we get the series of t elements in t minutes. For autoregressive model of order 1 (AR(1)), we need to calculate the ϕ to make the prediction by the following formula that is derived from Least Squares Method [4]:

$$\phi = \frac{\sum_{t=2}^{n}(X_t - \bar{X})(X_{t-1} - \bar{X})}{\sum_{t=2}^{n}(X_{t-1} - \bar{X})^2}$$

where

- ϕ: the autoregressive coefficient,
- ϕ_2: the autoregressive coefficient for AR(2) model,
- θ: method of moments estimate,
- X_t: the value of the time series at time t,
- $\bar{X}$: the mean of the time series data, calculated as

$$\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$$

.

For autoregressive model of order 2 (AR(2)), we calculate the ϕ_2 in the following way using Least Squares Method [4]:

$$\phi_2 = \frac{\sum_{t=1}^{n}(X_t X_{t-2}) - \phi_1 \sum_{t=1}^{n}(X_{t-1}X_{t-2})}{\sum_{t=1}^{n} X_{t-2}^2}.$$

For autoregressive moving-average model, we calculate the θ by minimizing sum of squared residuals [4]:

$$\theta = \frac{\sum_{t=2}^{n} \epsilon_t \epsilon_{t-1}}{\sum_{t=2}^{n} \epsilon_{t-1}^2}$$

where

- θ: method of moments estimate,
- ϵ_t error term in time t.

$$\epsilon_t = Y_t - \hat{Y}_t$$

and substitute the already found values in ARMA(1,1) model defined below:

$$Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1}.$$

We do the same for AR(1) model that is defined as:

$$Y_t = \phi Y_{t-1} + \epsilon_t$$

And for AR(2) model that has general form of:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t$$

Given the results obtained from above-mentioned models, we visualize the outcomes using graphical tools in Python and compare them to find the most appropriate results to the real prices.

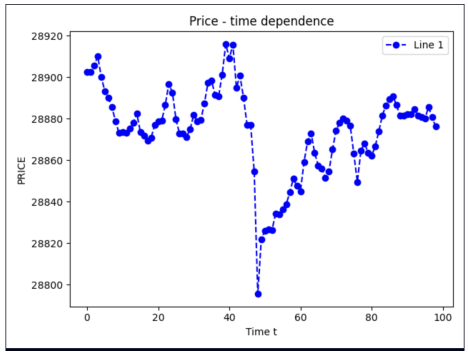

Figure 1: True price to time dependence

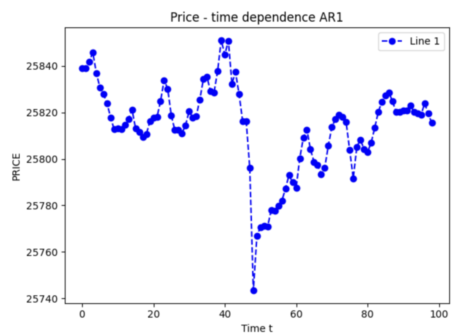

Figure 2: Price to time dependence - AR(1) model

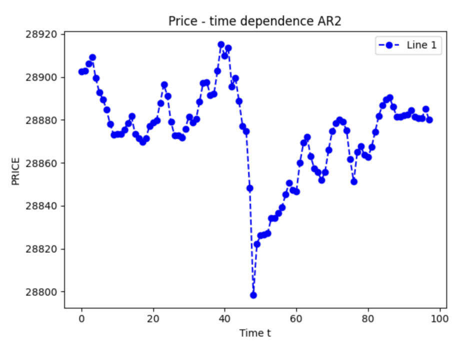

Figure 3: Price to time dependence - AR(2) model

3 Results

As we can see from the Figures 1, 2, 3, 4 we have presented all three models: AR(1), AR(2) and ARMA(1,1) gave correct predictions of bitcoin prices in

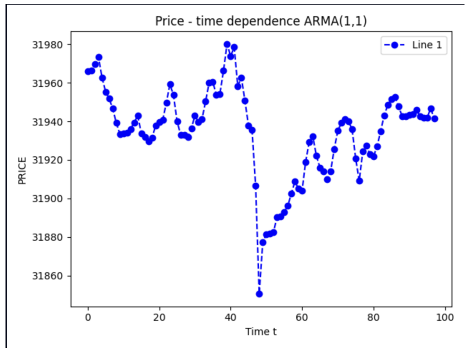

Figure 4: Price to time dependence - ARMA(1,1) model

Figure 5: Price to time dependence - AR(1) model

Figure 6: Price to time dependence - AR(2) model

range of approximately 11% for AR(1) model, 1% for AR(2) model and 10% for ARMA(1,1). Meaning of AR(2) was the closest to the real prices. However, according to the shapes of our graphs, all three models kept the same dependence

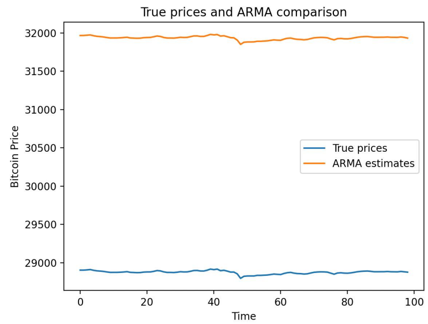

Figure 7: Price to time dependence - ARMA(1,1) model

of price and time t, which proves that all of them can be helpful in bitcoin price prediction. AR(1) being conservative, AR(2) being realistic and ARMA(1,1) being optimistic prediction.

In the graphs 5, 6, 7, we plotted model and true price graphs on the same axis

Using above mentioned formulas we calculated our results are:

- $\phi = 0.89400655102195$
- $\phi_2 = 0.10599540368363568$
- $\theta = 1.0000092440290393$

4 Comparison with existing models

As for the topic of keen interest, a lot of research has been done regarding cryptocurrencies and their prices forecasting. The comprehensive literature review has convinced us of the relevance of our research and provided insights on existing models, their nuances, and strengths. For instance, the research study conducted by Roy *et al.* [7] showed that Autoregressive-Moving Average model showed the highest accuracy (90.31%) in their study and so outperformed the AR and MA models because ARMA is more robust in comparison with other ones. Their accuracy was mainly evaluated using Root Mean Square Error (RMSE). Although ARMA seems to be the most suitable model for stationary time series predicting, Azari (2019) suggests that it may show errors in highly-volatile and non-stationary periods [2]. The author claims that ARMA tends to be more accurate for short-term predictions in sub-periods with consistent trends which can be improved by integrating additional features for more complex scenarios. Other approaches, including Long Short-Term Memory model (LSTM) [8], [9] or Generalized Autoregressive Conditional Heteroskedasticity model (GARCH) [5], [3] may also be productive in analyzing Bitcoin prices. Building upon these insights, our research sought to explore whether simpler models like AR(1) and AR(2) could exceed or complement ARMA under specific scenarios.

5 Conclusions

Cryptocurrencies like Bitcoin hold the interest of investors and financial analysts, making the topic worth researching by investigating models to foresee prices as accurately as possible. The choice among these models can be affected by current geopolitical, social and economic factors, such as wars, media trends, introduction of new technologies in industries, founding of new economic sectors and break down of others. Nevertheless, we tried to find the pattern to base our prediction on. Our study demonstrates that our modeling method can be used to determine accurate prices for bitcoin. AR(2) proved to be the most effective in terms of showing the most realistic result, however AR(1) and ARMA(1,1) respectively conservative and optimistic predictions can be helpful for an investor.

References

[1] Nashirah Abu Bakar and Sofian Rosbi. Autoregressive integrated moving average (arima) model for forecasting cryptocurrency exchange rate in high volatility environment: A new insight of bitcoin transaction. *International Journal of Advanced Engineering Research and Science*, 4(11):130–137, 2017.

[2] Amin Azari. Bitcoin price prediction: An arima approach. *arXiv preprint arXiv:1904.05315*, 2019.

[3] Zidi Gao, Yiwen He, and Ercan Engin Kuruoglu. A hybrid model integrating lstm and garch for bitcoin price prediction. In *2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6. IEEE, 2021.

[4] Kung-Sik Chan Jonathan D. Cryer. *Time Series Analysis with Applications in R*, volume 2. 2008.

[5] Quang Phung Duy, Oanh Nguyen Thi, Phuong Hao Le Thi, Hai Duong Pham Hoang, Khanh Linh Luong, and Kim Ngan Nguyen Thi. Estimating and forecasting bitcoin daily prices using arima-garch models. *Business Analyst Journal*, 45(1):11–23, 2024.

[6] M Poongodi, V Vijayakumar, and Naveen Chilamkurti. Bitcoin price prediction using arima model. *International Journal of Internet Technology and Secured Transactions*, 10(4):396–406, 2020.

[7] Shaily Roy, Samiha Nanjiba, and Amitabha Chakrabarty. Bitcoin price forecasting using time series analysis. In *2018 21st international conference of computer and information technology (ICCIT)*, pages 1–5. IEEE, 2018.

[8] Arif Mu'amar Wahid et al. Time series analysis of bitcoin prices using arima and lstm for trend prediction. *Journal of Digital Market and Digital Currency*, 1(1):60–74, 2024.

[9] Chih-Hung Wu, Chih-Chiang Lu, Yu-Feng Ma, and Ruei-Shan Lu. A new forecasting framework for bitcoin price with lstm. In *2018 IEEE international conference on data mining workshops (ICDMW)*, pages 168–175. IEEE, 2018.