

Price Prediction of Bitcoin, Ethereum and XRP Based on the ARMA Model

Ziyan Zhang

University of Bristol, Bristol BS8 1TH, UK

fs20022@bristol.ac.uk

Abstract. The growing potential and high volatility of the cryptocurrency market attract a lot of interest from both businesses and investors. Even though the prices fluctuate, predicting with time series models such as ARMA and ARIMA would still provide a useful reference for analyzing the market. Recent studies on machine learning methods including RNNs have made new progress in forecasting digital currencies. This study focuses on one of the traditional models ARMA to predict the time series dataset from 2021-2022 of cryptocurrencies including Bitcoin, Ethereum and Ripple. To be specific, AIC and ADF tests are used to choose the optimal model and suitable dataset. According to the analysis, the ARMA model would be affected by the volatility of Bitcoin. However, the predictions are not precise enough but still a valuable reference for certain businesses and individual investors. More state-of-art machine learning models can be utilized in future study to enhance the performance. Overall, these results shed light on guiding further exploration of cryptocurrency price prediction.

Keywords: ARMA; Price prediction; Bitcoin; Ethereum; XRP.

1. Introduction

Cryptocurrency price predictions have attracted a lot of investment interest in the recent past. In the business world, cryptocurrencies have significantly increased the possibilities of the financial and commercial markets. Forecasting models help bitcoin investors select the best investments and could result in higher profits and returns. Cryptocurrencies are digital or virtual currencies designed to transfer digital money online. Transactions generated on the decentralized platform are protected by encryption technologies through steps like verification on a blockchain and sending digitally signed messages. One of the most popular cryptocurrencies is Bitcoin, which was the first contemporary cryptocurrency released in 2009. An anonymous group established this contemporary cryptocurrency under the fake name Satoshi Nakamoto, and it quickly gained popularity all over the world. It is still used today as the most used and valued digital currency in the world. The first strategy and protocol for Bitcoin were outlined forth in Nakamoto's 2008 article. In January 2009, Bitcoin was released as open-source software on SourceForge [1].

Since the entire source code is open to the public, it is easy for people to make changes and produce a new cryptocurrency. This kind of digital currency is called alt-coins since they are an alternative to Bitcoin. Later digital cryptocurrencies made use of similar core technology as Bitcoin. As a result of the public's interest in Satoshi Nakamoto's work, everyone else has created alternative cryptocurrencies to meet their specific needs.

Various cryptocurrencies and blockchains that compete with it or are seen to have limitations have inspired altcoins to work to overcome such issues. Later, whole new currencies were created and consensus rules were reconsidered, efficiency was increased. The efficiency was also increased due to the introduction of smart contracts. In 2011, the first alternative currency Litecoin was removed from the Bitcoin network. Another cryptocurrency is Ether (ETH) which hasn't separated from Bitcoin. It was developed by Vitalik Buterin, Dr Gavin Wood, and other developers in support of Ethereum, the biggest blockchain-based virtual machine in the world. Ether (ETH) is used to compensate network users for the work done on their computers to validate transactions. The expansion of digital venture capital investment and the advancement of distributed ledger technology are the two key factors promoting the industry's expansion. Businesses are using the benefits of

blockchain technology and digital currency to invest in cryptocurrencies and work together to deliver effective and high-quality services to their clients.

At the time of writing, there are over 20,000 cryptocurrencies in circulation, with a combined value of cryptos at over \$2.2 trillion which is 7% of the narrow money supply of the world [2]. Thus, determining the future projections of the cryptocurrency in the overall currency market is a current topic for researchers that may yield valuable information on cryptocurrency market share. Probabilistic modelling is used to analyze people's investment choices, including buy and sell orders and field studies on sentiment or peer influence [3].

Recent research has shown that social media data can be utilized to track changes in investor sentiment and the price of Bitcoin and other well-known cryptocurrencies [4]. Dartmouth researchers recently employed Gradient Boosting Tree Model to produce sentiment-based predictions on changes in the price of alternative cryptocurrencies by examining Twitter posts' sentiment and the number of transactions made [5]. The findings suggest that an extreme gradient tree regression model may be able to predict price swings in the cryptocurrency market.

Traditional models like Capital Asset Pricing Model (CAPM) are easy to use but have drawbacks including using inputs that may cause volatility [6]. A lot of research has been done on the use of machine learning techniques to estimate cryptocurrency prices in the last few years. Classic models, including ARMA and ARIMA, standing for Auto Regressive Moving Average and Auto regressive Integrated Moving Average models, respectively, as well as Recurrent neural networks (RNNs) which have gained increasing popularity recently, can use a wide range of endogenous and exogenous inputs to generate prediction results. However, RNNs struggle with the issue of vanishing gradients [7]. The parameter updates lose significance when the gradient decreases too much. Long data sequences are therefore hard to learn.

The combination of AR and MA models is the ARMA model. It has more accurate estimates than the AR and MA models. The ARMA model has two orders of p and q and is called the ARMA (p , q) model. Compared to ARMA, ARIMA could difference the non-stationary variables until it achieves stationary. In contrast to the two orders of ARMA (p , q), ARIMA is a triplet of orders (p , d , q) called the ARIMA (p , d , q) model. ARMA and ARIMA will act precisely the same if the data is already stationary [8].

Compared to the described traditional models and machine learning methods, another model known as the Fama and French model has been tested in numerous global financial markets [9]. In an attempt to overcome the drawbacks of the CAPM model, Fama and French presented their three-factor model. Fama French model evaluates daily returns and common fluctuations in a more efficient way while CAMP is suitable for investigating the sensitivity of the general market.

Motivated by the above discussion, this paper concentrates on analyzing the prediction of the traditional model ARMA in the price trend of the crypto markets, with the performance of the Fama and French model on selected cryptocurrencies Bitcoin (BTC), Ethereum (ETH) and Ripple (XRP). The trend of the currency is not only related to historical data but also to many external factors, such as the attention of users and the policies of different countries. However, these external factors do not need to consider in this study. The only factor that uses in this study is the digital coin's historical prices and the ARMA time series models use to predict the coin's trend.

2. Data & Method

2.1 Data

The datasets used in the study are extracted from finance.yahoo.com. Data from 8th August 2021 to 8th September 2022 of three top-rated cryptocurrencies, BTC, ETH and XRP are taken into account for prediction. The price of cryptocurrencies fluctuated dramatically at these chosen times, unlike the typical financial markets. Each dataset row includes following features: Date, Currency, Open, High, Low, Close, and Adjusted close price, where close price is selected to be the dependent variable. It is

also important to take adjusted close price into account at some point since it is modified for corporate actions and often used when examining historical prices.

2.2 ARMA Model & AIC

The full name of AR is Auto Regressive. AR stands for Auto Regressive. It combines white noise with a linear combination of past points and can predict future points. In the context of data mining, white noise can be thought of as an arbitrary process with an expectation of zero and a constant variance. The AR model also has an order called the AR (p) model, also known as the p-order autoregressive model. It refers to the linear combination of the previous p points and white noise to forecast future points.

One different feature of the Moving Average (MA) from the AR model is that it is a linear combination of data from a previous time series set. It considers past white noise to affect the current point in time; in an AR model, the historical white noise indirectly influences the forecast at the current point in time by influencing the historical time series values. Similarly, the MA model has an order known as the MA (q) model or the qth order moving average model. Generally, ARMA refers to ARMA (p, q) model containing p terms of the AR model and q terms of the MA model where the formula is:

$$X_t = c + \epsilon_t + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (1)$$

Here, ϕ is auto regressive parameter, θ is moving average parameter, c is constant, and ϵ is the white noise. Since one doesn't know what p and q are optimal for the model, one can set an interval range for them, for example, range (0, 3), calculate the AIC values of the different models, and choose the ARMA model with the smallest AIC value. This optimal ARMA model is then used to predict the average price with the selected time scale of BTC, ETH and XRP over the next 2 years and visualize the results. Models like ARMA and ARIMA cannot forecast on non-stationary time series data, so the initial stage for time series forecasting is to determine how many times of differencing are needed to turn the series stationary. ADF test is used to check the stationarity here.

According to the different time series (day, month, quarter, year), one can compress the data and visualize them. Month is selected as the timescale of the prediction model. The average price of bitcoins for that month is calculated first in the datasets. After compressing the raw data in months, one can get the average price of bitcoins for that month. The selection of the model's parameters p, q, and r is one of the most important steps in the estimate process. This is important since it helps to obtain the best possible forecast. One of the methods is graphical, employing the plot of ACF and PACF. The other method is the Akaike information criterion, which is selected for the model due to its automation. For example, one can calculate the marginal likelihood at any time. Akaike information criterion is used to check if the how well a statistical model fits. The smaller the AIC value, the better the model fits [10]. The equation is:

$$AIC = -2\ln L + 2k \quad (2)$$

Here, K is number of variables used and log-likelihood is used to test how well the model fits. The higher the number, the better the fit. The AIC will rank each model from best to worst. It is unable to know the optimal model for p and q, so one can set an interval for them, e.g. range (0, 3) and then calculate the AIC values of the different models and choose the ARMA model with the smallest AIC value. This optimal ARMA model is then used to predict the future average price of the currency.

Other criteria have also been considered. BIC (Bayesian Information Criterion) is an alternative to AIC which would however take into account the number of rows of the dataset. The model performs better on a selected dataset if the BIC value is lower. Both AIC and BIC are used to help find more reliable results. HQIC (Hannan–Quinn information criterion) can also be used for feature extraction.

3. Results & Discussion

3.1 Feature Extraction

Previous research attempted to lower the variations by utilizing alternative time intervals, such as hourly and daily while comparing the mean values of bitcoin datasets and found considerable differences. On the other hand, selecting a particular time series frequency (such as daily, weekly, or monthly) can produce various observations. Monthly price is chosen in this research data, which reduces the data's dimensionality and saves the ARMA model's training time. The Fig. 1, Fig. 2 and Fig. 3 represent the price of BTC, ETH and XRP in the day, month, quarter and year. Seen from the results, the trend of BTC, ETH and XRP by month and the trend by day are similar. Using month as the time scale could reduce the local fluctuations and reflect the tendency more efficiently simultaneously, thus reducing the model's training time.

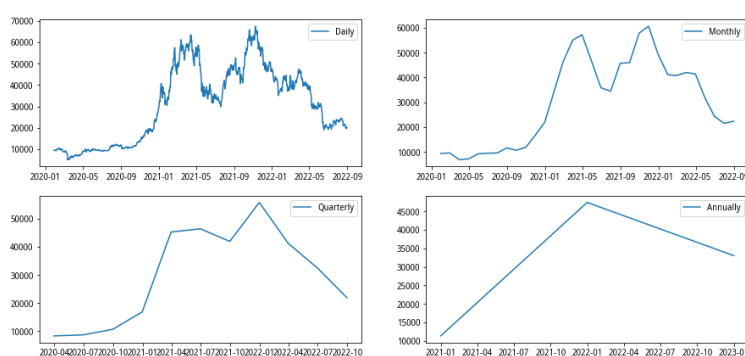


Fig. 1 Prices of BTC.

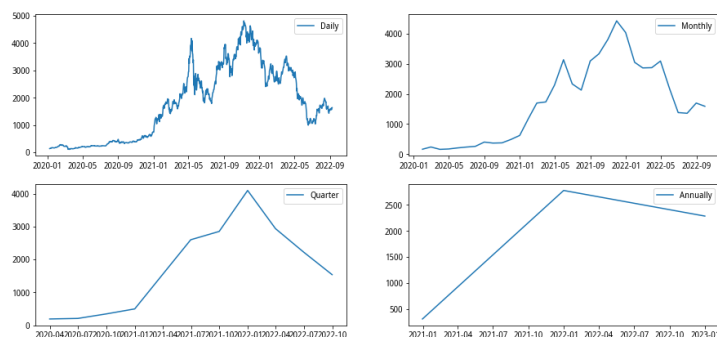


Fig. 2 Prices of ETH.

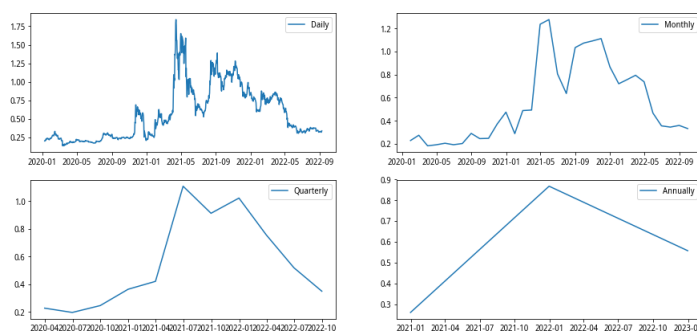


Fig. 3 Prices of XRP.

3.2 ADF

The Augmented Dickey-Fuller Test (ADF) is a test used in time series analysis to identify the presence of unit roots in the sample dataset and check if the data is statistically significant (stationary). Any time series analysis may produce unexpected outcomes if unit roots are employed. The

hypotheses for the test are listed as H_0 : There is a unit root; H_1 : The dataset is stationary. The P-value of a stationary dataset is usually close to 0. According to the ADF test, the dataset can be also considered stationary if the test results are less than the critical value at 1%, 5%, or 10% significance, H_0 (the null hypothesis) will be rejected. The more negative the test result is, the stronger the rejection of H_0 at one of 1%, 5%, or 10% significance, indicating that there is a unit root exists. The t value is -1.42, -11.89, -7.01 for BTC, ETH and XRP, respectively. The test results of close prices show that t-statistic values of ETH and XRP are both smaller than the critical values specified in the test at 1%, 5% and 10% levels, showing that the time series data is stationary. Since the ARMA model requires stationarity, the selected dataset is relatively reliable to be used for training. However, the t-statistic of BTC is greater than the 10% value, so first difference is needed to achieve stationary. Figure 4 shows the result after the first difference.



Fig. 4 BTC results after the first difference.

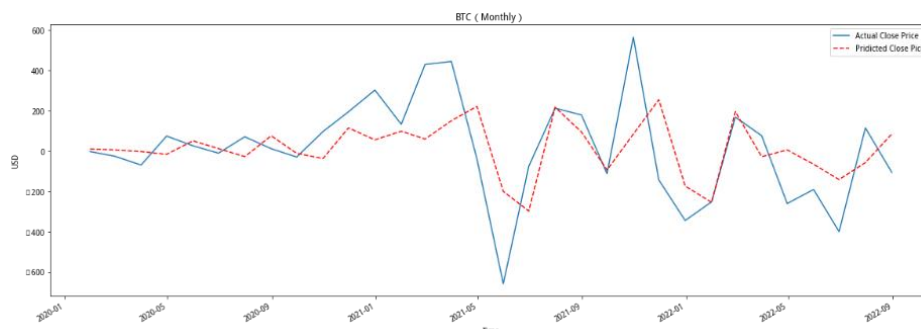
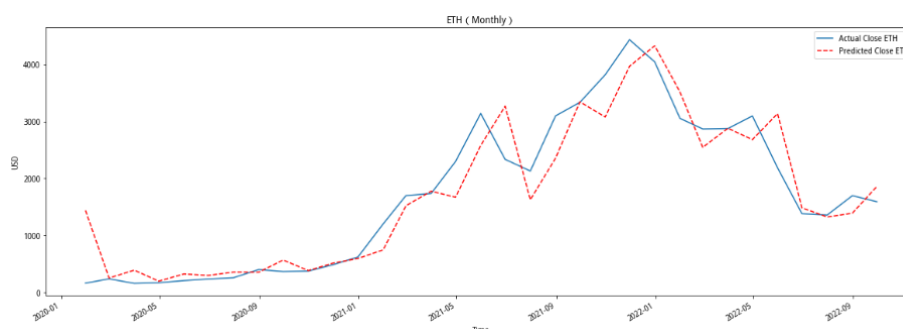
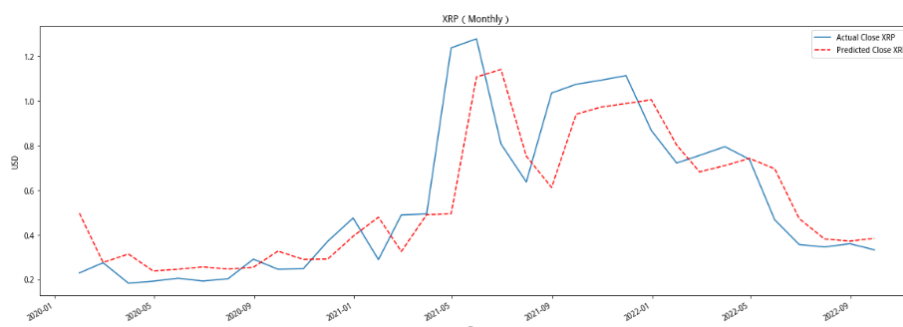
For the ARMA prediction, p and q should be specified. The model parameters with the smallest AIC values were saved by creating all possible combinations of (p, q) in the range $(0, 3)$ and calculating the AIC values for each ARMA (p, q) model. The optimal model has been calculated for the selected cryptocurrency datasets. The next stage of the analysis is to create ARMA (close (p, q)) model. The data is predicted for the period 2021-2022 then plot results. According to the results displayed in Table. 1, the number of observations is reduced to less than 100. The log-likelihood take log of historical data and it is a representation of how the model fits. Recent researches prefer log-likelihood since it is a more straightforward way to show the maximum likelihood. This value is relatively meaningless if the evaluated model is less than one since comparisons are needed. Generally, the higher the log-likelihood, the better the model fits.

Table 1. The optimal models and metrics.

	Optimal model	Log-likelihood	AIC	BIC	HQIC
BTC	ARMA (2, 0)	-321.21	650.41	656.28	652.38
ETH	ARMA (1, 1)	-246.06	500.11	506.10	502.13
XRP	ARMA (1, 0)	7.64	-9.29	-4.80	-7.78

3.3 Empirical Analysis

The outcomes of applying ARMA model to BTC, ETH and XRP are illustrated in Fig. 5, Fig. 6, and Fig. 7, respectively. The price was plotted against time (monthly), where the actual cryptocurrency price is shown in the blue line and the prediction price is shown in the red dot line. Different from other cryptocurrencies, bitcoin's volatility will lead to uncertain future predictions. The actually recorded bitcoin price reached a much higher or lower point than the model predicts while the date the outliers occurred was much similar. As time series models tend to take up some regularities in patterns over time, producing misleading results, further reducing trends and seasonality improves forecasts.

**Fig. 5** Price prediction of BTC.**Fig. 6** Price prediction of ETH.**Fig. 7** Price prediction of XRP.

Due to the massive market valuation of cryptocurrencies, predicting their prices is a research field with great financial benefit potential. The goal of this study was to justify the usefulness of the ARMA model of price prediction for the three most popular cryptocurrencies. When predicting a specific dataset, regression analysis can be used if the relationship between multiple variables and the outcome is being considered, or time series analysis can be used to consider the relationship between a single time dimension and the outcome.

4. Conclusion

In summary, ARMA model performs better on cryptocurrency with lower volatility. In addition to historical data, the trend of the currency is also related to many external factors, such as the interest of users, the policies of various countries, and the market sentiment. Although the overall market value of cryptocurrencies is rising, making them more related and similar to traditional currencies, these digital currencies still face problems such as security and volatility. Despite the above factors, the results of the ARMA model could still be used as a reference to help some specific business to predict the future trend of digital currencies and provide useful information. Overall, these results offer a guideline for cryptocurrency price prediction based on time-series model.

References

- [1] Coelho, D. M. C.: Application of the Fama French 3-Factor. Lisbon, (2020).
- [2] REIFF, NATHAN. How Much of All Money Is in Bitcoin? 26 11 2021. <<https://www.investopedia.com/tech/how-much-worlds-money-bitcoin/>>. (2021).
- [3] Peter, M. K., Nicolas, D. P., and Alex S. P.: An Experimental Study of Cryptocurrency Market Dynamics. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). Association for Computing Machinery. New York, (2018).
- [4] Gorse, R. C., Phillips, D.: Predicting cryptocurrency price bubbles using social media data and epidemic modelling. 2017 IEEE symposium series on computational intelligence (SSCI) (pp. 1-7). IEEE (2017).
- [5] Li, T. R., Chamrajnagar, A. S., Fong, X. R., Rizik, N. R., Fu, F.: Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model. Frontiers in Physics, 7, 98 (2019).
- [6] Zucchi, K.: CAPM Model: Advantages and Disadvantages. Prieiga per internetą: < <https://www.investopedia.com/articles/investing/021015/advantages-anddisadvantages-capm-model>. Asp, (2019).
- [7] Sherstinsky, A.: Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena, 404, 132306, (2020).
- [8] Wirawan, I. M., Widiyaningtyas, T., Hasan, M. M.: Short term prediction on bitcoin price using ARIMA method. In 2019 International Seminar on Application for Technology of Information and Communication (iSemantic) (pp. 260-265). IEEE (2019).
- [9] Kapur, V.: Is the Fama and French model a good indicator of market sectoral performance? Study of the relationship between excess industry returns and the. New York, (2007).
- [10] Gilli, M., Maringer, D., Schumann, E.: Numerical methods and optimization in finance. Academic Press (2019).