Bitcoin Price Prediction

Alina Armijo

Hunter College CUNY

2021

Table of Contents

Introduction

Bitcoin is a cryptocurrency invented in 2008 by an unknown person or group of people using the name Satoshi Nakamoto. The currency began use in 2009 when its implementation was released as open-source software.

Bitcoin is a decentralized digital currency, without a central bank or single administrator, that can be sent from user to user on the peer-to-peer bitcoin network without the need for intermediaries. Transactions are verified by network nodes through cryptography and recorded in a public distributed ledger called a blockchain. Bitcoins are created as a reward for a process known as mining. They can be exchanged for other currencies, products, and services.

Bitcoin has been criticized for its use in illegal transactions, the large amount of electricity used by miners, price volatility, and thefts from exchanges. Some economists, including several Nobel laureates, have characterized it as a speculative bubble at various times. Bitcoin has also been used as an investment, although several regulatory agencies have issued investor alerts about bitcoin.

As of January 2021, there are more almost eight thousand cryptocurrencies in the market. Among those, the most famous and outstanding one is Bitcoin. Its Market Capitalization, which is calculated by multiplying bitcoin's price by its current circulating supply, is 1,041.38 billion U.S. dollars (USD) as of February 19, 2021. It occupies 62.5% of the market compared to the second popular cryptocurrency, Ethereum, which takes only around 10% of the total Market Capitalization. One reason that Bitcoins rise to fame is that it can be exchanged and spent conveniently and globally with a low transaction fee. Users can buy and sell bitcoins through both online exchanges and online bitcoin ATMs. Many people purchase Bitcoin as an investment, which makes the prediction of Bitcoin attractive. A successful forecast of the future price or the price trend can provide abundant profits since investors can use such information to assist in decision-making.

This study aims to find a suitable model to predict the closing price of bitcoin.

Dataset

2.1 Data Overview

The raw data set is taken from kaggle.com and includes the 1-minute interval Bitcoin information for January 2012- December 2020. It includes eight variables, shown in Table 2.1, and 4,727,777 observations that are in chronological order. Each observation includes price information and volume information for every 1-minute time step.

Table 2.1 Variables Description

Variables	Meaning
Open	Price of the first trade at the current time step
High	Highest Price of trades at the current time step
Low	Lowest Price of trades at the current time step
Close	Price of the last trade before the next time step
Volume_BTC	Trade volume in Bitcoins
Volume_Currency	Trade volume in USD
Weighted_Price	Weighted Bitcoin price
Timestamp	Ten digit integer that represents time
	*All prices are in U.S. dollars (USD)

Before working with the data we need to clean it. There are 1,243,472 missing values in the raw dataset. It could be possible that there was no trade within 60 seconds, which led to the empty record of such information in the dataset. The data was missing for a long time due to the Bitstamp Exchange reporting a hack and losing less than 19 thousand coins in January 2015. The exchange suspended the service from January 05 to January 09, 2015, for investigation, which is the longest termination since the establishment of the Bitstamp Exchange. Since the missing observations do not provide much useful information, and the rest of the data is sufficiently large, we eliminate NA values.

2.2 Exploratory Data Analysis

We have seen that Bitcoin has become extremely popular. Now let's take a close look on the price change with time. From Table 2.2 we can see that its closing price rose from 1.5 dollars to 19,666 dollars. We also see that the price change is highly volatile with a standard deviation of 4104.8 dollars. For better visualization of my data I constructed a sample using systematic selection. Since raw data contains 4,727,777 observations with 60-second intervals, I decided that the subset of about 300000 observations which will cover about 100 observations a day will be sufficient enough to work with.

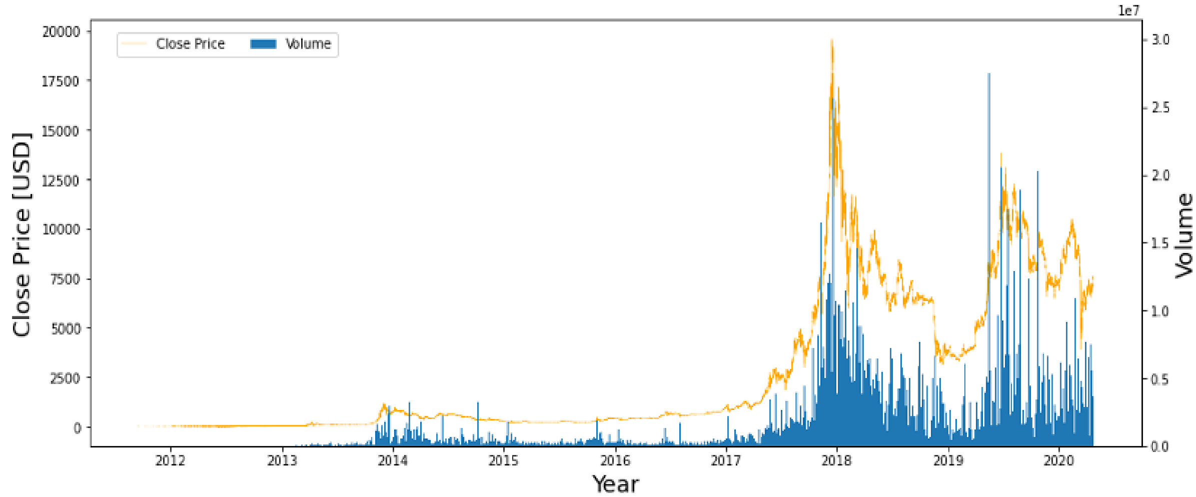
Table 2.2 Raw Dataset Distribution

Variables	Min	Max	Mean	Standard Deviation	Median
Open (USD)	3.8	19665.8	4056.1	4104.9	2389.3
High(USD)	3.8	19666.0	4058.9	4108.2	2391.0
Low(USD)	1.5	19650.0	4053.1	4101.3	2386.3
Close(USD)	1.5	19665.8	4056.0	4104.8	2389.3
Volume_BTC	0.0	5853.9	9.6	31.61	1.94
Volume_Currency	0.0	10445888	30653	107030.9	2937
Weighted_Price (USD)	3.8	19663.3	4056.0	4104.8	2389.0

Now, let's see when fluctuations in price have occurred. According to Figure 2.1 at the early stage price was very low and pretty stable. A huge price jump happened at the end of 2017 when the price rose to nearly 20,000 dollars which is about 13000 times of its minimum of all the times, then it dropped back to about 8,000 dollars. After the second peak in the middle of

2018, the price dropped again to a very low price of about 3,000 dollars. The price increased again to about 13,000 dollars in the summer of 2019.

Figure 2.1: Price and Volume of Bitcoin



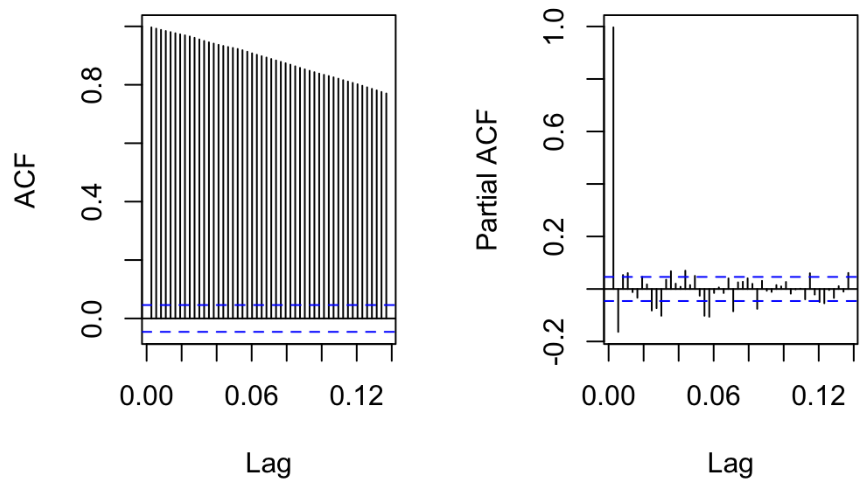
It is very important to see the change in Volume_BTC as well. The Bitcoin trading volume indicates how many Bitcoins are being bought and sold on specific exchanges. High trading volumes are likely to drive more on-chain activity, for example when people deposit and withdraw funds. It is also a good indicator of the general interest in the crypto market. It would be logical to think that the volume of Bitcoin is associated with its price. Going back to Figure 2.1 we see that there are a few transactions that appeared when the price was below 1,000 dollars. Then the volume reached its peak in 2018 when price was at the peak as well.

Besides closing price and volume, other variables also play a crucial role in the forecast of Bitcoin. The opening price and closing price over the same time period are strongly correlated. If we take the difference between closing and opening price over the same period it will be close to zero and symmetric along zero. It indicates that the Bitcoin price increases or decreases about the same amount.

2.3. Data Preprocessing

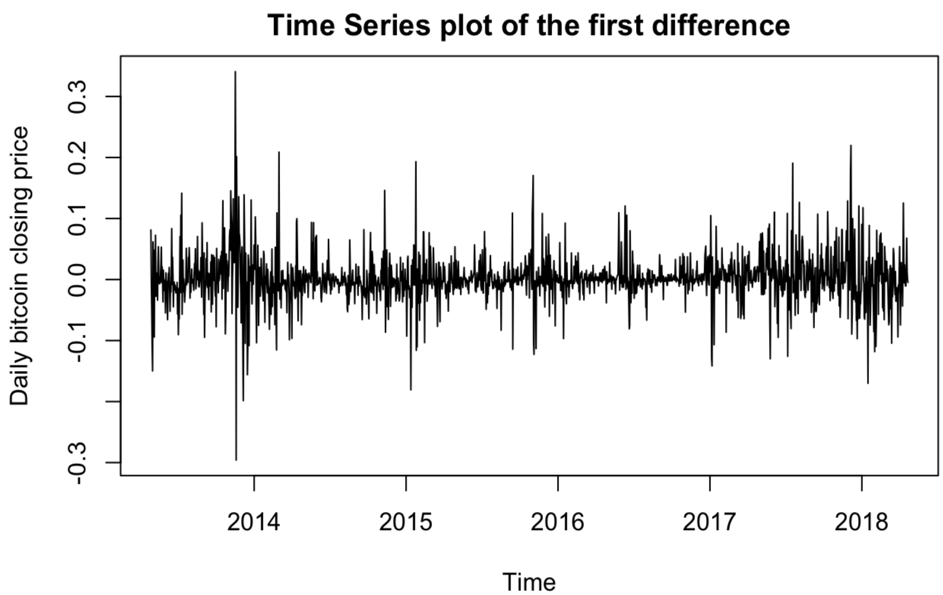
The ACF of the data on Figure 2.2 decays gradually, whereas the PACF cuts off abruptly. The gradual decay of the ACF tells us that we should difference the data, which would also address the issue of the upward trend. Also we are going to use log transformation of the data.

Figure 2.2 ACF and PACF



Since one of the assumptions when using the ARIMA model is having constant variance, differencing will help us with that. We will take the first difference between one period of time(t) and the period of time before that (t-1). Figure 2.3 shows that there is no obvious trend in the time series plot after first differencing. However there is a huge volatility in price change of Bitcoin in the end of 2013 and 2017 years. While Bitcoin was achieving recognition on the international level, some countries were banning trading on bitcoin exchanges. For example, on 23 June 2013, it was reported that the US Drug Enforcement Administration listed 11.02 bitcoins as a seized asset in a United States Department of Justice seizure notice pursuant to 21 U.S.C. § 881. This marked the first time a government agency claimed to have seized bitcoin. This and some other events such accepting bitcoin as payment for tuition fees lead to a big price increase. However, in December 2013, Overstock.com announced plans to accept bitcoin in the second half of 2014. On 5 December 2013, the People's Bank of China prohibited Chinese financial institutions from using bitcoins. After the announcement, the value of bitcoins dropped, and Baidu no longer accepted bitcoins for certain services. This explains volatility in the end of 2013 year. In 2017 the number of businesses accepting bitcoin continued to increase. On 6 December 2017 the software marketplace Steam announced that it would no longer accept bitcoin as payment for its products, citing slow transactions speeds, price volatility, and high fees for transactions [Ref2]. These events are possible explanations for high volatility in 2017-2018 years. Therefore, when modeling we will assume constant volatility but keep in mind that there was times when this assumption is violated due to events described above.

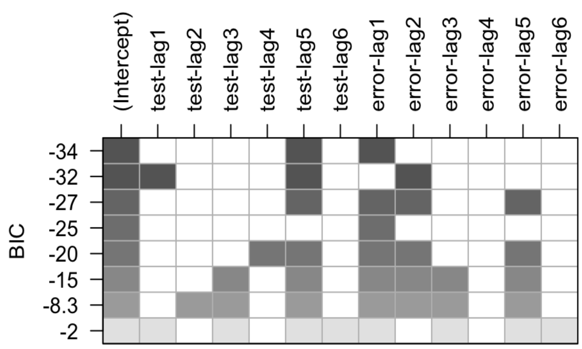
Figure 2.3



Implementation

3.1 ARIMA Model Selection

From Figure 3.4 we see the set of possible models from EACF is {ARIMA(1,1,1), ARIMA(2,1,1), ARIMA(2,1,4), ARIMA(4,1,4), ARIMA(2,1,2)}



AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	o	o	o	x	o	o	o	o	x	o	o	o	o
1	x	x	o	o	x	o	o	o	o	x	o	o	o	o
2	o	x	x	o	x	o	x	o	o	x	o	o	o	o
3	o	x	x	o	x	x	o	o	o	x	o	o	o	o
4	x	x	x	x	x	o	o	o	o	x	o	o	o	o
5	x	x	x	x	x	o	o	o	o	x	o	o	o	o
6	o	o	x	x	x	o	o	o	o	x	o	o	o	o
7	o	x	o	o	x	x	o	o	o	x	o	o	o	o

Figure 3.4 BIC and EACF tables of transformed data

3.2 ARIMA Modeling

After comparing different models from my set of possible models I decided that ARIMA(2,1,1) model is the best one for the given data. We can see the summary of this model on Figure 3.5. This model has one of the lowest AIC which is -6616.89, Sigma squared is 0.001535 which is pretty much the same in all diagnosed models.

Figure 3.5 ARIMA(2,1,1)

```
Call:
arima(x = diff.bitcoin.t, order = c(2, 1, 1), method = "ML")

Coefficients:
         ar1      ar2      ma1
      0.1486  -0.0695  -0.9921
s.e.  0.0252   0.0252   0.0108

sigma^2 estimated as 0.001535:  log likelihood = 3311.45,  aic = -6616.89
```

So we have the ARIMA(2,1,1) model

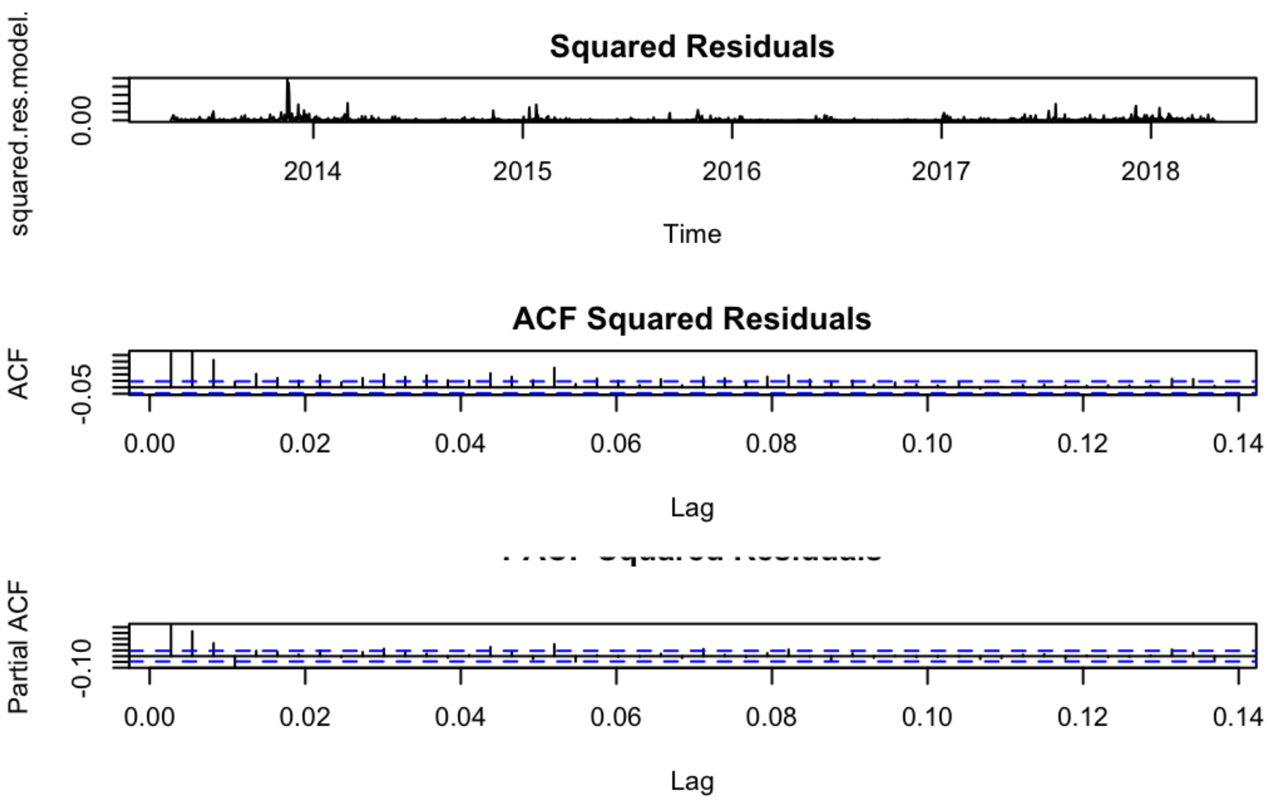
$$X_t - 0.1486X_{t-1} + 0.0682X_{t-2} = Z_t + 0.9921Z_{t-1}$$

where Z_t should be approximately white noise with mean 0 and variance σ_Z^2.

3.3 Diagnostics of the Model

The model that fits the data should be tested to determine the quality of the fit. Figure 3.6 shows Residuals Analysis for ARIMA(2,1 ,1). Squared Residuals plot shows clusters of volatility at the same points in time specific events have occurred as described in Section 2.3.

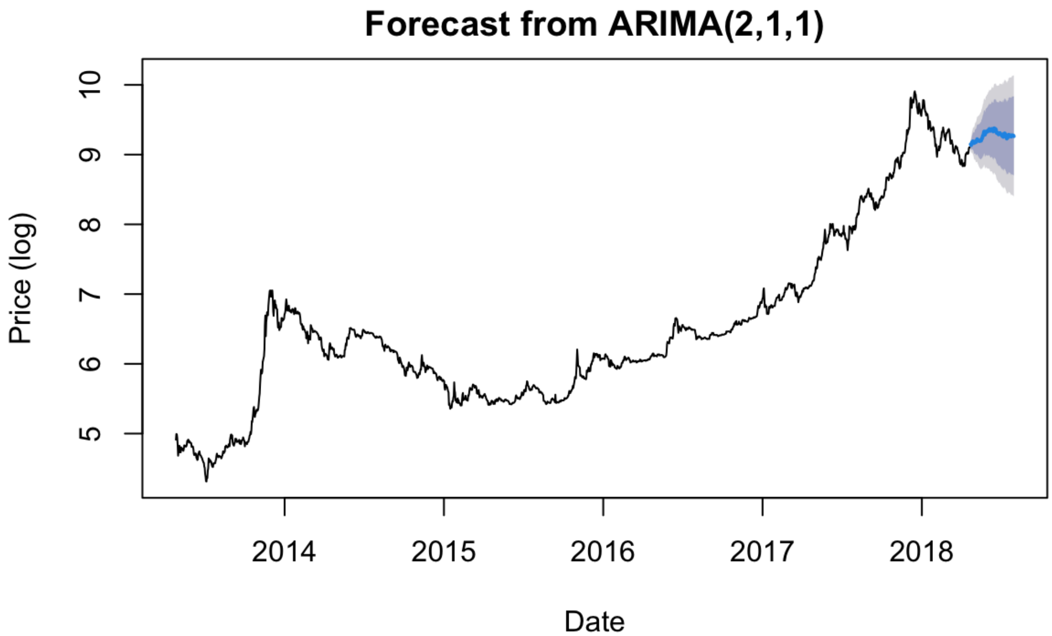
Figure 3.6 Residual Analysis



3.4 Forecasting

Figure 3.7 shows forecasting of ARIMA(2,1,1) with 80% and 95% prediction intervals. We do observe that the confidence intervals of the predictions capture the true values. The further one goes into the future, the larger the confidence interval i.e. the more uncertain predictions become. To evaluate the performance of the prediction I used the mean absolute percentage error (MAPE), which is 5.88%. Based on this MAPE value I conclude that this model shows a high degree of reliability in predicting the price of Bitcoin.

Figure 3.7



Conclusion

The goal of this project was to find a suitable model to predict the closing price of bitcoin. The process included exploratory analysis, data visualization, model specification, model fitting and selection, followed by diagnostic checking and forecasting.

Ultimately, our two lowest AICc models produced predictions that were relatively close to the true data, with the ARIMA(2,1,1) model. Even though this model did not really pass model diagnostics it's the best fitted model for the given data.

The biggest challenge for me was to preprocess the data and select the sample to be analysed. This dataset was difficult to fit with an ARIMA model because of the fact that the variance was non-constant, which results in squared residuals that were dependent on one another.

Looking at the historical price of Bitcoin we noticed big volatility in price due to certain events occurring. Therefore it is extremely hard to make any predictions. The residuals were also not Normal, even after transforming with the Box-Cox transformation in the beginning.

If I would continue working to accurately forecast on this dataset, I would try to utilize a GARCH model instead, which better takes into account the non-constant variance. GARCH is a typical model for financial data.

References

[Kag1] Data is taken from kaggle website

https://www.kaggle.com/mczielinski/bitcoin-historical-data

[Bit1] Financial Innovation volume6, Article number: 13 (2020)

https://jfin-swufe.springeropen.com/articles/10.1186/s40854-020-0174-9

[Ref1] https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average

[Ref2] https://en.wikipedia.org/wiki/History_of_bitcoin#2013