

Introduction

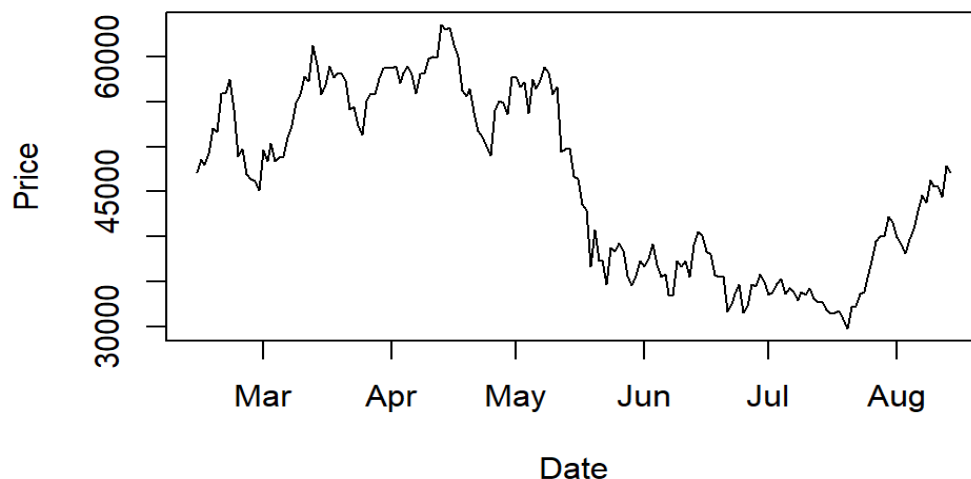
Bitcoin is a decentralized virtual currency, which is also known as digital cryptocurrency. As a different form of "online money" than physical money, it is used to transfer funds via network transmission. Over the past few years, Bitcoin has gained in stature among consumers, governments, and financial traders. Bitcoin is now the largest digital cryptocurrency in the world, not only in terms of market capitalization, but also in terms of its popularity and appeal.

Bitcoin as an investment product with high risk and high return, its price has had several big ups and downs in 2021, so for this project, I decided to explore the trend of Bitcoin in the short term. Will the price of Bitcoin increase or decrease? If the Bitcoin is overvalued, then we can consider selling it now at a high price; if the Bitcoin is undervalued, then we can consider buying it at a lower price.

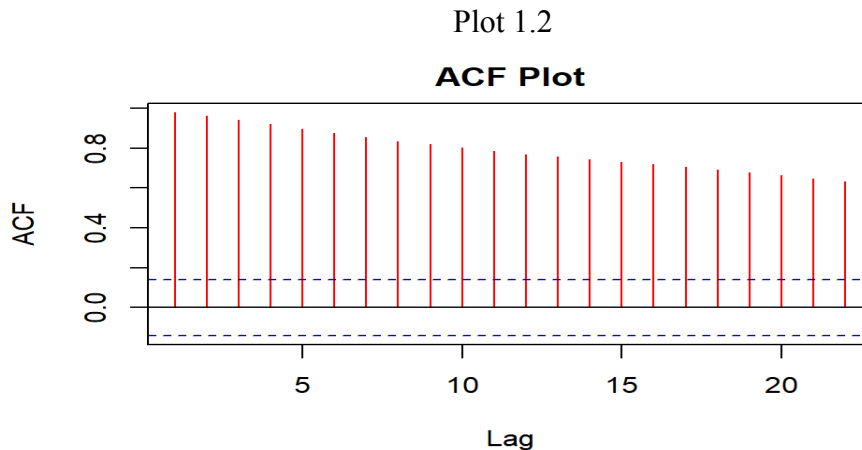
Modelling

Based on the plot below (1.1) of Bitcoin prices every day from the date 2021-2-13 to the date 2021-08-24, we can see there are many fluctuations, and the change of price was volatile over 183 days. Although this plot has already obviously shown to be non-stationary, we still ought to check the ACF plot of Bitcoin price. ought to check the ACF plot of Bitcoin price.

Plot 1.1



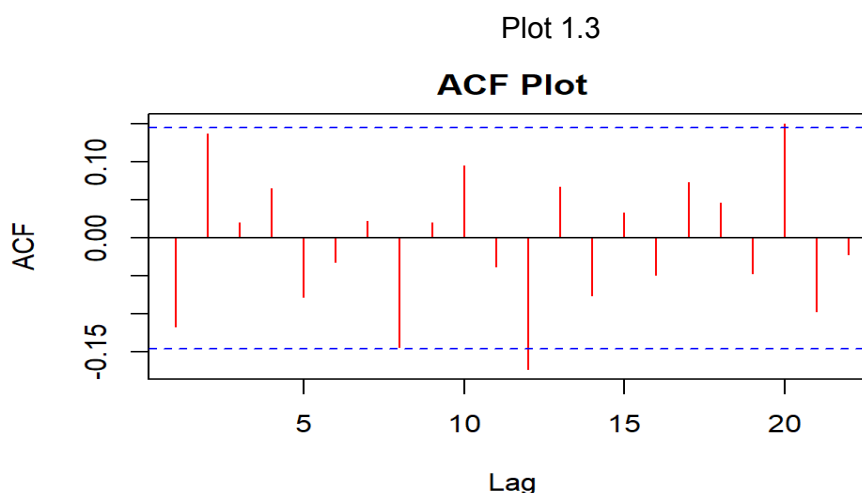
From the ACF plot below (1.2), we observe that the plot decays slowly, meaning that the time series is nonstationary. The future prices depend highly on the past prices, and no seasonality present as well.



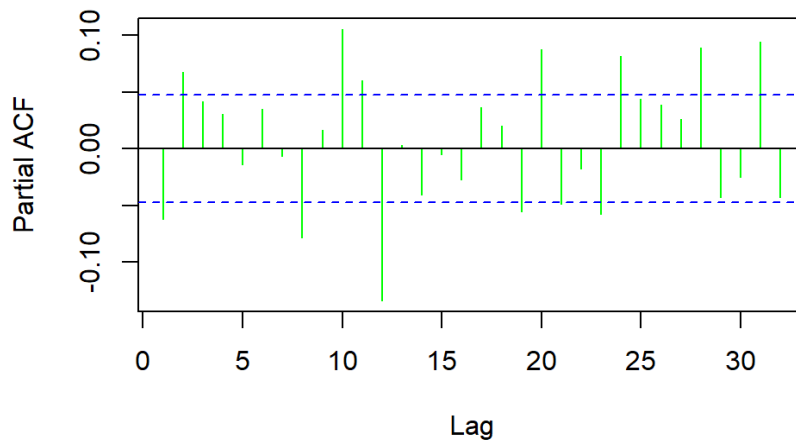
Looking at the basic statistics of the time series of Bitcoin prices, we observe that the mean value is 46000.3 which is non-zero, and the variance 93039322 is significantly high. This indicates again that the time series is not stationary. Thus, to attain a stationary time series, we need to analyze the difference of Bitcoin price.

Based on the difference of Bitcoin price, we would argue that the time series becomes relatively stationary since it has an overall mean -0.4791209 which is approximately to 0, and no obvious pattern or seasonal effect observed.

Since after differencing, the time series is relatively stationary, then we can consider which model is most suitable for this time series. Firstly, from the ACF plot (1.3), it is obvious that after lag 2 there is a rapid cutoff, so we may consider it to be an IMA(2) model. Then, from the PACF plot (1.4), when lag=3, the value of PACF is in the range. So we conclude that it also might be an IAR(2,1) model. In addition to that, we should also consider the combination of ARI(2,1) and IMA(1,2), which means the ARIMA(2,1,2) is also a consideration.



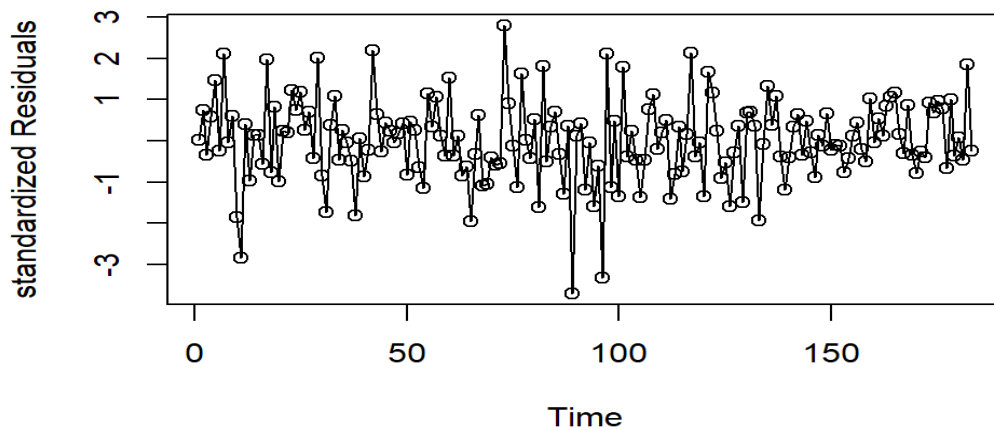
Plot 1.4



In order to select the optimal model, we are required to find their Akaike information criterion (AIC) for comparison. By calculation, the AIC of ARIMA(2,1,2,) is equal to 3274.11, the AIC of IMA(1,2) is 3274.22, and the AIC of ARI(2,1) is 3274.42. Since ARIMA(2,1,2) has the smallest value of AIC among these models, we prefer to select ARIMA(2,1,2) for further diagnosing and prediction. Then, we can calculate the estimate parameters for the model by using maximum likelihood estimation, and get the result the coefficients of ar1, ar2, ma1, ma2 are 1.1565, -0.8081, -1.2186, 0.9475 respectively.

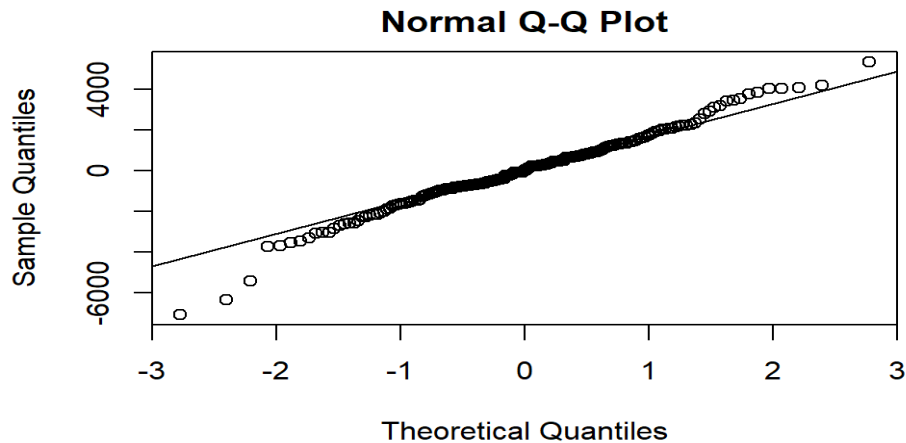
Plot (1.5) displays the time series plot of the standardized residuals from the ARIMA(2,1,2) model estimated for the difference of the Bitcoin price time series. There are at least 2 residuals early in the series with magnitudes larger than 3, which is unusual in a standard normal distribution. But as a whole, most of the residuals are scattered randomly along the axis $y=0$, so we can conclude the residuals have constant variance.

Plot 1.5



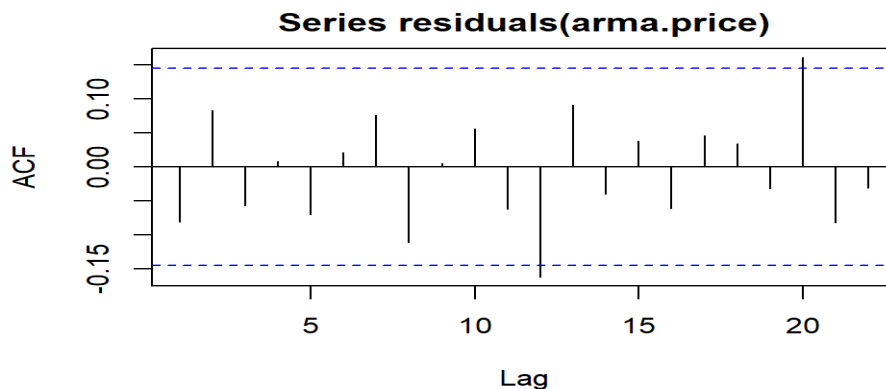
The quantile-quantile plot for the residuals from ARIMA(2,1,2) model for the difference of Bitcoin prices time series is displayed below (1.6). Obviously, most of the points are located along the straight line, but on the two tails, there are still 4 outliers. Still, we can conclude the residuals follow normal distribution.

Plot 1.6



A graph of the sample ACF of these residuals is shown in plot (1.7). The dashed horizontal lines plotted are based on the large lag standard error of $\pm 2/\sqrt{n}$. Only lag 12 and lag 20 exceed the horizontal lines a little, we can ignore this. So there is no evidence of autocorrelation in the residuals of this model.

Plot 1.7

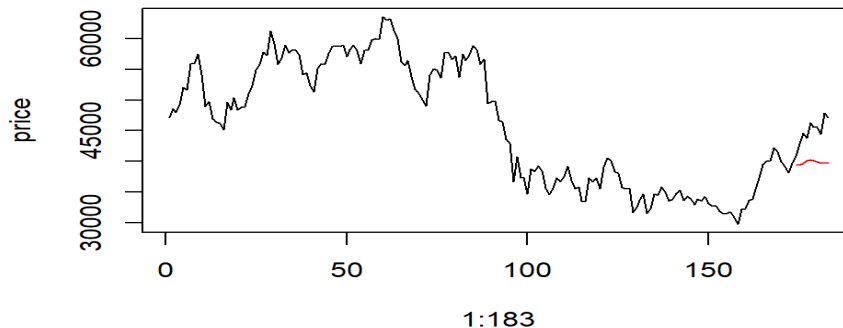


Results:

Let us predict the last 10 days (from 2021-08-04 to 2021-08-14) Bitcoin Price by using our ARIMA(2,1,2) model. We retain the last 10 observations for forecasting and use the first 173 observations to fit the models. And the predicted Bitcoin price for the last 10 days are 39349.00, 39389.03, 39732.50, 40094.34, 40244.31, 40138.01, 39901.40, 39712.56, 39678.32, 39784.1.

To have a more direct comparison, we can see the original data and our prediction in the following plot (1.8), the black line is shown the historical data, and the red line is our prediction.

Plot 1.8



The purpose of this project is to forecast the change of Bitcoin price, but by the prediction, it shows the price will not have a huge increase or decrease. My prediction indicates that the price will increase around 2000 in the next five days. Based on my prediction, the Bitcoin price is undervalued at the date 2021-08-03, so buying it on that day is a considerable decision and selling it five days later. By comparing with the actual data, we can see that the growth rate for the next five days is actually much higher than our expectation. So we can conclude that on this day (2021-08-03) Bitcoin is undervalued and it is a good time to buy in. To some extent, the trends we predicted were accurate.

Conclusion:

Although in this project, we predicted the trend of Bitcoin price in the next ten days by ARIMA(2,1,2) model, our predicted trend is similar to the actual situation. In our prediction Bitcoin will go up by almost 2000 in the next 10 days, but it actually went up by 12,000, which is much larger than our prediction. However, there still exist limitations. Since we cannot be sure if the price of bitcoin is a time series because it is influenced by many other factors such as politics and economy. If it is not a time series, then we can't use a time series model to predict it. In addition to that, the supply and market demand for bitcoin also affects the price of bitcoin. This is because its price is non-regulated and is determined by supply and demand. If the Bitcoin protocol allows the creation of new bitcoins at a fixed rate. Then the demand for bitcoin grows faster than the supply increases, thus pushing up the price of bitcoin. Because the price of Bitcoin receives many complex external factors, we cannot tell whether it is a set of time series or not.

Bitcoin's price has risen from a few cents to tens of thousands of dollars now, and his price ups and downs have been extreme in the investment market. But based on the first two reductions in bitcoin production, as bitcoin halves and production decreases, bitcoin will become more and more scarce, and after that will usher in a super bull market, and as the price of bitcoin soars, long-term holders of bitcoin will be the biggest beneficiaries.

References:

1. Adari, Umesh Kumar. "Historical Bitcoin Data." Kaggle, August 24, 2021. <https://www.kaggle.com/datasets/umeshkumar017/historical-bitcoin-data>.
2. "Sign In." RPubS. Accessed April 8, 2022. https://rpubs.com/riazakhan94/arima_with_example.
3. "Sign In." RPubS. Accessed April 8, 2022. https://rpubs.com/RRD27/ts2_gen.
4. Reiff, Nathan. "Why Bitcoin Price Predictions Are Unreliable." Investopedia. Investopedia, February 8, 2022. <https://www.investopedia.com/tech/why-bitcoin-price-predictions-are-unreliable/>.