Dogecoin to the ground

A forecasting of the Dogecoin price

Siming Shan

Abstract

Elon Musk, as one of the influcential men in the world, not only has huge impact on technology, but also has a huge impact on the fianancial market. Last year, Elon Musk bought a huge amount of Dogecoin(A cryptocurrency), and highly recommended it on his twitter by saying that Dogecoin is people's cryptocurrency, and because of that the price of Doogecoin soared up in a dramaticly speed last year. Hwever, one year later, is Dogecoin still worth inveseting? In this paper, we focus on building an ARIMA model to fit the Dogecoin price data, and then predict the price for the next ten days. We found that the Dogecoin price fits an ARIMA(1,1,1) model, and by using this model we predict that the price of Dogecin will keep decreasing for the next ten days.

keywords: Elon Musk, Bogecoin, ARIMA, Finance, Prediction, Cryptocurrency

Contents

1 Introduction:

Elon Musk is known as one of the richest men in the world, and the CEO of the Tesla Motors. On March 14th, Elon Musk proposed to buy the whole twitter company with price of 43 billion dollars has brought him to the public attention again, however,as a matter of fact, Elon Musk has bought 9% stake of twitter and this news caused the twitter shares soared more than 20% []. It is not the first time Elon Musk has had a huge impact on the finance market.

last year was a big year of crypto currency because Elon Musk greatly advertise a cryto currency that he bought - Dogecoin. Dogecoin is a open-source cryptocurrency created by Billy Markus and Jackson Palmer[]. The cryptocurrency is a digital or virtual currency which is secured by crypotography[]. In 2018, as Satoshi Nakamoto created the first cryptocurrency bitcoin in the world, the concept of decentralized networks transaction has been more and more accepted and populated around the world. Nowasays, the crypotocurrency investment has become the first choice of many investors, because it is secure, decentralized, and stable.

On Elon Musk twitter, he claimed that the "dogecoin is the people's crypto", and on May 10th 2021, he claimed that SpaceX will launch satellite called "Doge-1" to the moon, thus "Doge to the moon" has become a famous meme in the crypto currency area. As a matter of fact, the doge coin did actually soar up after Elon Musk's tweets. However, after the tide of dogecoin has passed, and Elon Musk decreased his frequency of tweeting about dogecoin, the price of dogecoin rapidly fell. Does the dogecoin still worthy to invest? This paper uses the statistical programing language R[] and fucous on using the auto-regressive integrated moving average(ARIMA) model to modeling the dogecoin price from 1st May 2021 to 23rd April 2022 to forecast the dogecion price in the next ten day. The data was collected from Yahoo finance by using the function `getSymbols` from the `quantmod` package[]. In the orginal dataset, there are 358 observations with six variables each, these variables are the open price, the high price, the low price, the close price, the volume, and the adjusted price. We focus on the close price since it represents the last price during a regular trading day.

ARIMA is a statistical analysis model used to forecast time series dataset such as the unemplyment rate, stock price, or the global temperature. The ARIMA can forest a future value Based on the past values and past errors, the mathematical equation of this model will be discussed in the method section. By fitting this model to oour data, We found that the price of dogecoin fits a ARIMA(2,1,1) model, and by forecating based on this model, the price of dogecoin after 23rd April 2022 will first slightly increase for one day, then decrease for nine days, the overall price of the dogecoin in the next ten days presents a downward trend.

2 Data

2.1 Time series data introduction

In time series analysis, the model created cannot be accurate if the sample size is too small, thus a dataset with over 100 observations is desired for modeling. Our dataset was collected from Yahoo finance, and it recorded the open price, the high price, the low price, the close price, the volume, and the adjusted price of dogecoin from 1st May 2021 to 23rd April 2022, and thus there are 358 observations. Since there are 358 observations in the dataset, we could build a relatively precise model if there are some underlying relationship between observations. The dogecoin was orignally released in 2013, however, we are not going to use all the observations for modeling, because the price of the dogecoin was so low for a period of time, as Figure 1 shown, which may largely influence the accuracy of the model, thus we focus on the price of dogecoin from 1st May 2021 to 23rd April 2022. Based on Figure 2, observed although all five different prices have different values, there trends and shapes are very similar, the prices were declining rapidly from around 0.7 USD to around only 0.2 USD after May 2021, after that the prices went up and down for few times but the overall trend was decreasing. So we only have to anlyze one of these price for modeling, in this case, we choose the closing price. Figure 3 shows the volume of the Dogecoin for the past 358 days, although volume could also give investors some insights of whether entering the market or not, in this paper will only focus on the price.

2.2 Data visualization

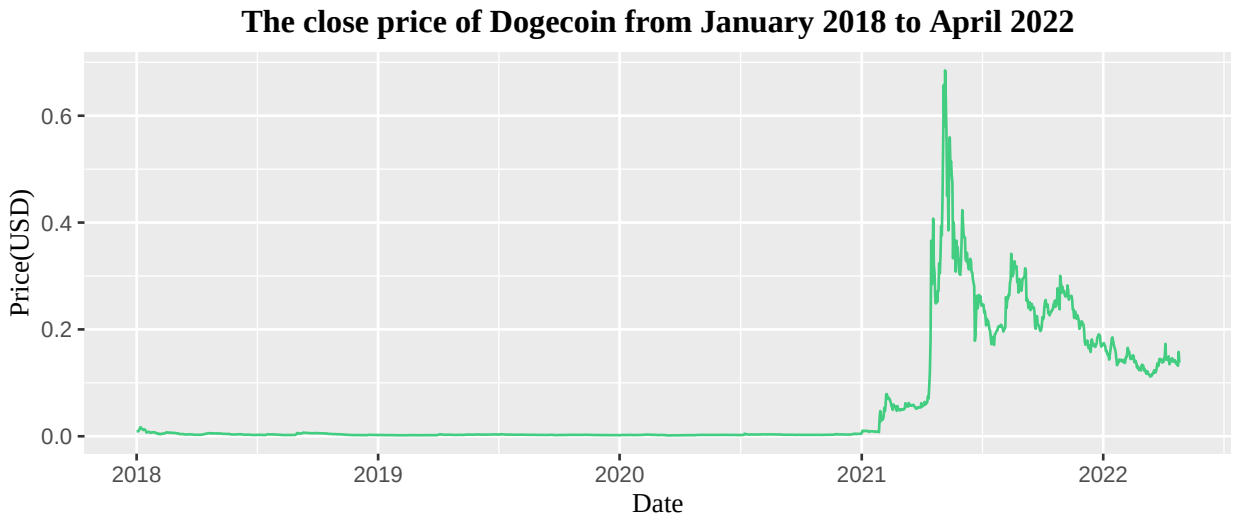

Figure 1: The close price of Dogecoin from 2018 to 2022

Figure 2: Five different prices of the Dogecoin from May 2021 to April 2022

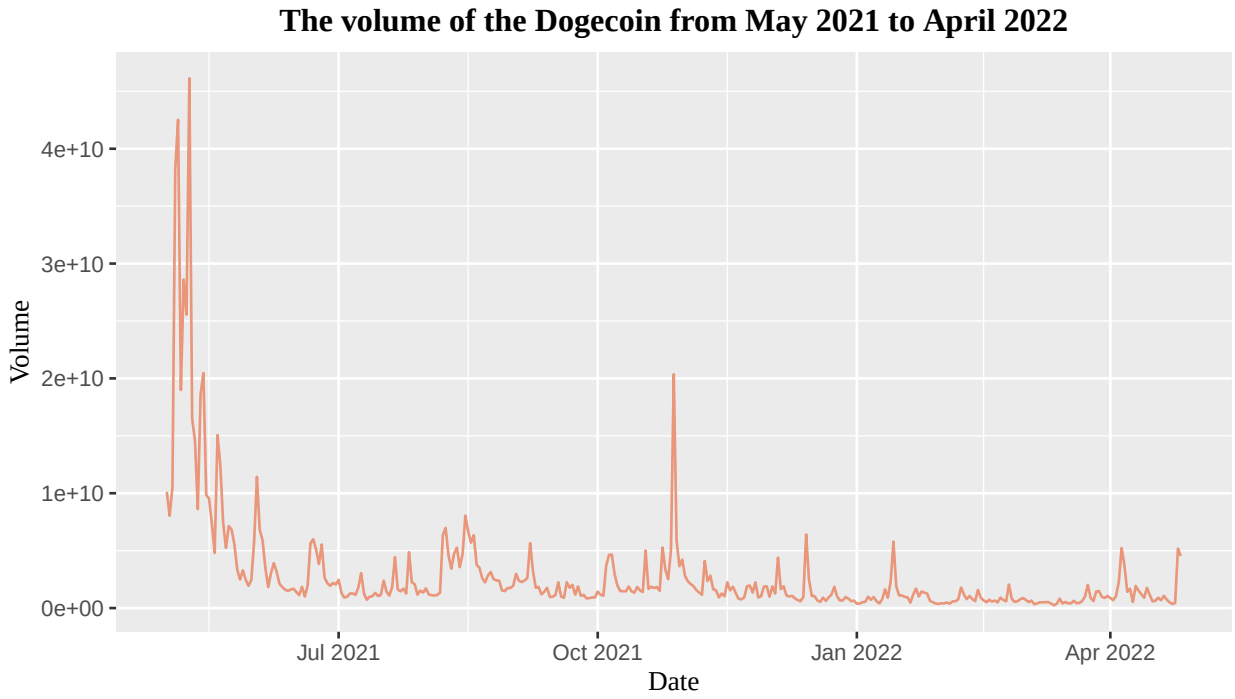

Figure 3: The volume of the Dogecoin from May 2021 to April 2022

3 Method

3.1 ARIMA model

Autoregressive integrated moving average model ARIMA(p,d,q) is one of the most common and useful model to predict future values based on past values and past errors, where p represents the autoregressive order, d is the dependent order(the number of difference transformations needed), and q iis the moving average order. The mathematical equation of a ARIMA model can be expressed as follows:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + ... + \phi_p X_{t-p} + \theta_0 + \theta_1 W_{t-1} + \theta_2 W_{t-2} + \theta_3 W_{t-3} + ... + \theta_q W_{t-q}$$

In the above equation, X_t is the value to predict, $X_{t-1}, X_{t-2}, X_{t-3}...X_{t-p}$ are past values, ϕ_p are the coefficients of the past values, $W_{t-1}, W_{t-2}, W_{t-3}...W_{t-p}$ are past white noise errors, θ_p are the coefficients of the past errors.One important assumption to build a ARIMA model is that the time series data is stationary with no seasonality. A stationary time series is defined as time series that have:

- Constant mean over time t
- Constant variance over time t
- The autocovariance function between two observations X_{t1} and X_{t2} only depends on the interval t1 and t2.

So before fitting model to the time series data, changing the data in to stationary without sesonality is essential, and there are generally two ways to do that. The first method is Applying difference transformation: transform each observations at time t into the difference between observation at time t and observation at time (t-1),the mathematical equation is defined as $\Delta X_t = X_t - X_{t-1}$, while the second method is applying logarithmic transformation, which takes log of each transformation.

Base on Figure 4, an obvious trend is observed, which means this time sereis is not stationary, thus a first difference transformation was applied on the data. The Figure 5 shows the price of the Dogecoin after the first difference transformation. We observed that the mean became constant over time, however, the variance is still obviously not constant over time since the plot shows a fanning pattern. Thus we applied a Box-Cox transformation to stablized the variance. The Figure 6 shows the price of the Dogecoin after both transformations, we observed that both mean and variance have been stablized, thus we would use this Box-Cox transformed data to fit a model. At this point, the dependent order of the ARIMA(p,d,q) model can be determined, which is d=1.

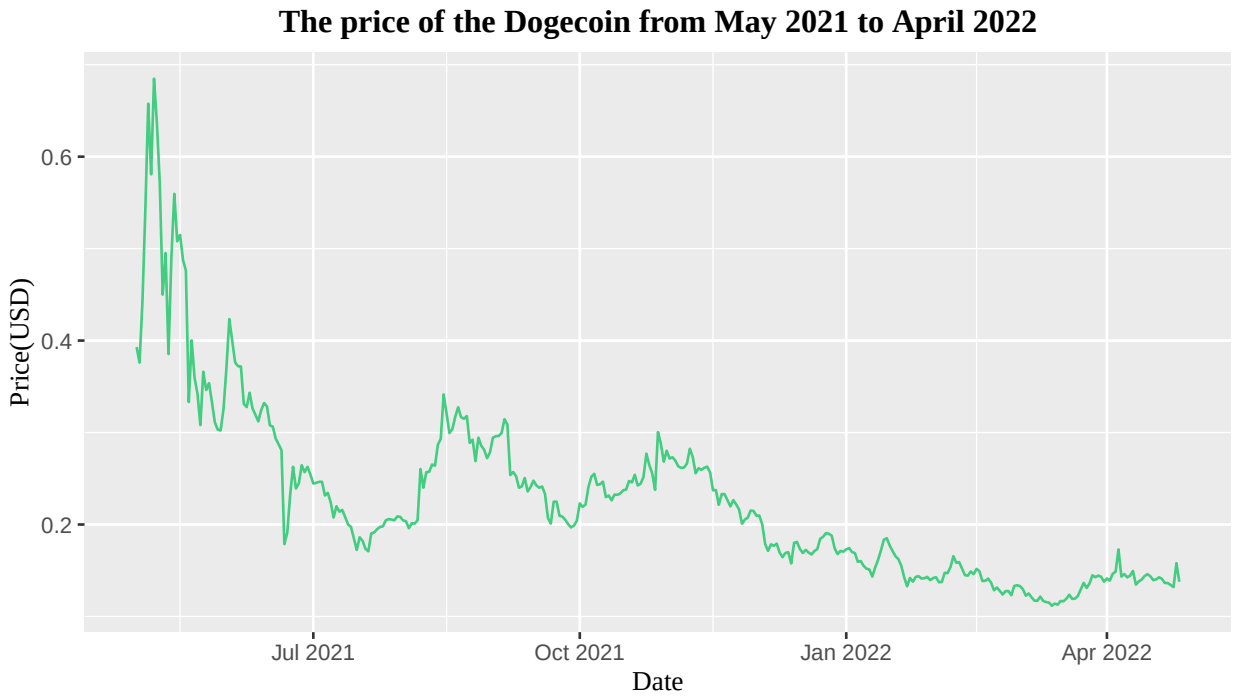

Figure 4: The price of the Dogecoin from May 2021 to April 2022

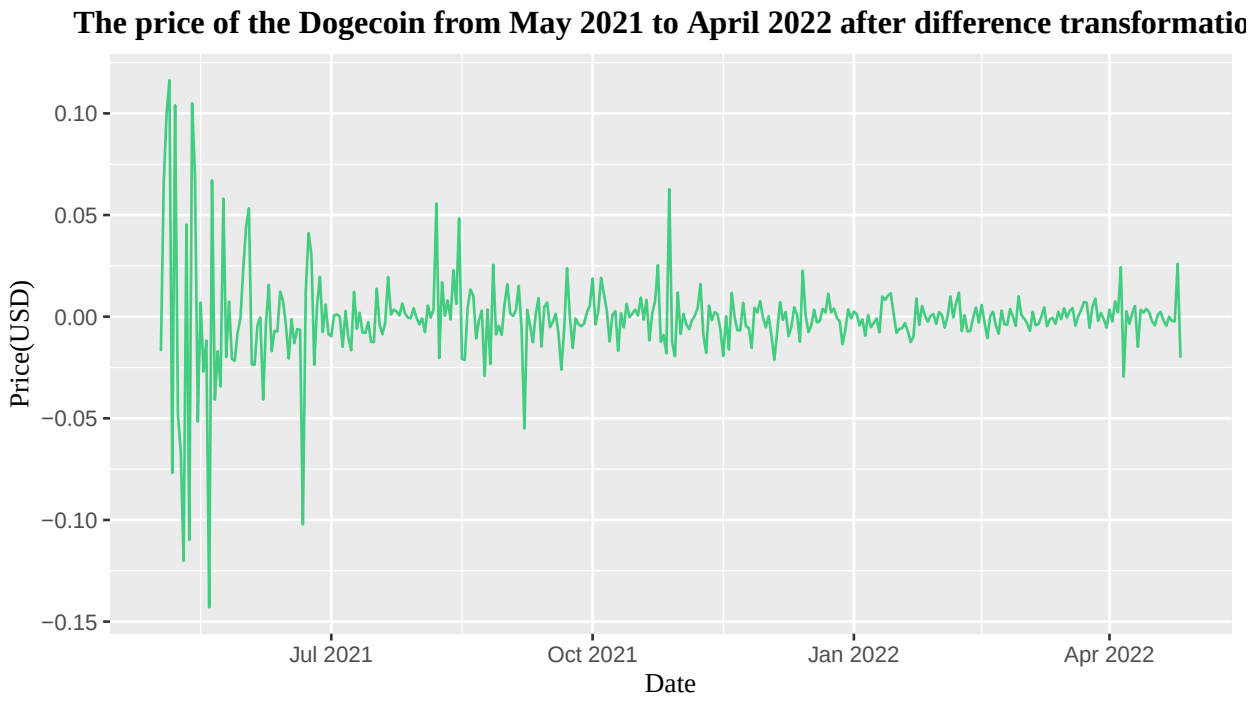

Figure 5: The price of the Dogecoin from May 2021 to April 2022 after difference transformation

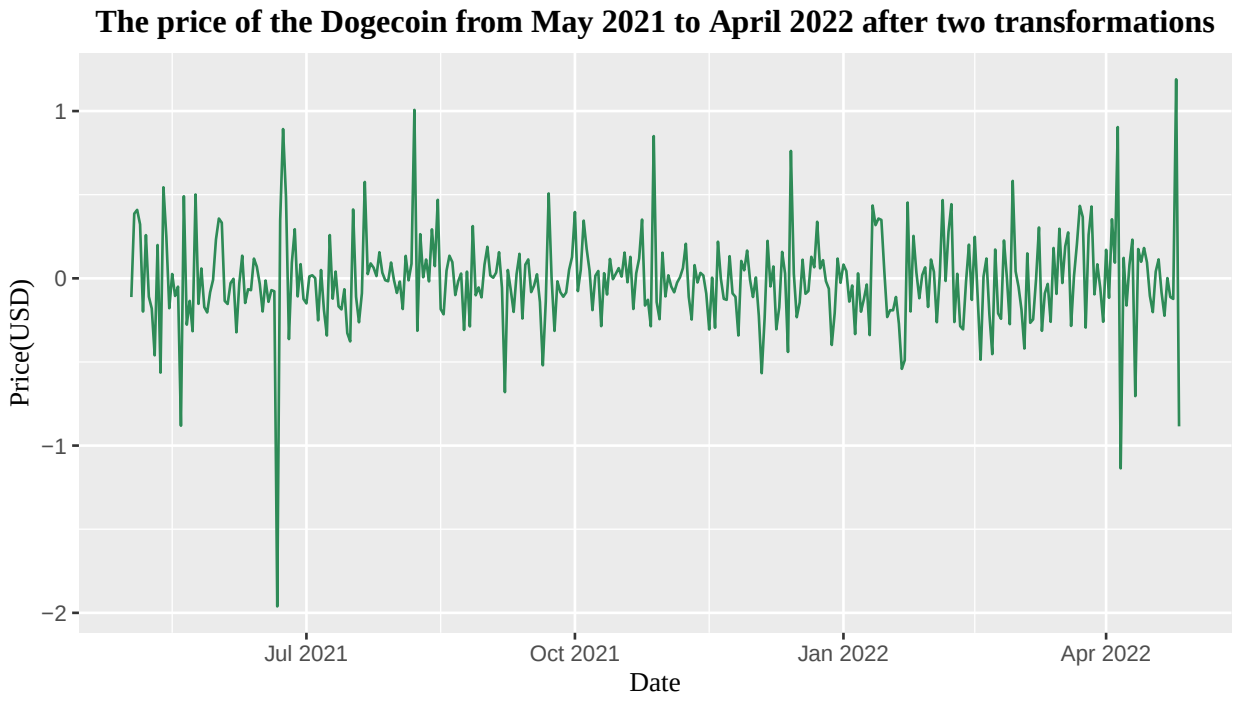

Figure 6: The price of the Dogecoin from May 2021 to April 2022 after difference and Box-Cox transformation

3.2 ACF and PACF

A Autocorrelation function (ACF) plot and a Paritial autocorrelation function (PACF) plot were used to determine the AR and MA order. Autocrrelation function calculate the correlation between one observation(X_t) with it's lag(X_{t-h}), while PACF also measure the correlation between X_t and its lag at time t X_{t-h}, in addition each correlation controls for any correlation between observations of a shorter lag length[]. Figure 7 shows the ACF and PACF plots of the transformed time series data. The ACF plots shows a gradually declining to zero after the first lag. The PACF also shows a gradual declining to zero aftter the first lag. So the AR order of the model could be 1 or 2 since PACF is also quite large at secnd lag, and the MA order of the model could be 1 since the ACF declining raplidly after lag 1. Thus two possible model can be introduced, ARIMA(1,1,1) and ARIMA(2,1,1)

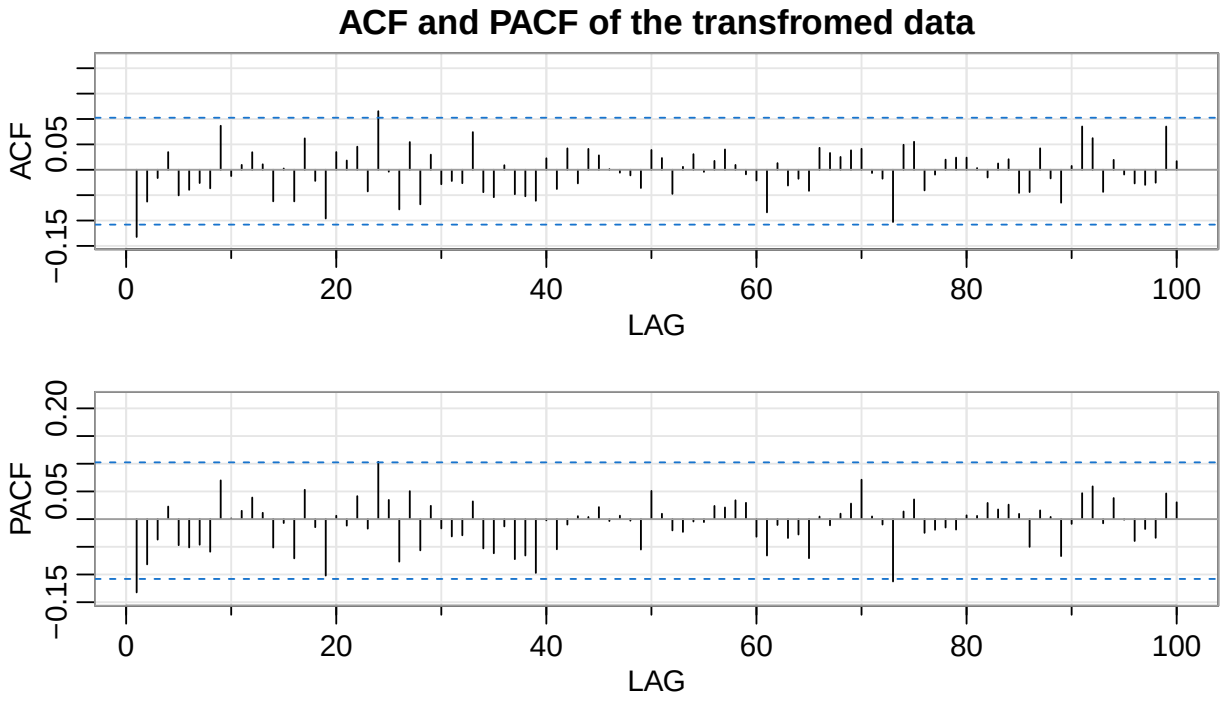

Figure 7: TACF and PACF of the transfromed data

3.3 Model diagnostics

For model diagnostics, as Figure 8 and Figure 9 shown, we plot a residual plot, a ACF of residual plot, a normal-QQ plot, and a Ljung-Box statistic plot to examine wheter the model has violated any model assumptions or not. Then we compare the Akaike information criterion (AIC) of each plots, finally, we compare the significance of their coefficients to find the best fit model.

3.3.1 Model assumptions

A good model should not violate any of the model assumptions like normality, independecy, and randomness. We can use a standard residual plot and ACF residual plot to determine wheter the model vioolate the randomness assumption. Use Normal Q-Q plot to determine whether the model violate the normal assumption, and use Ljung-Box statistics plot to determine whether the model violate the independent assumption. Figure 8 and Figure 9 shows the necessary model assumptions of two models.

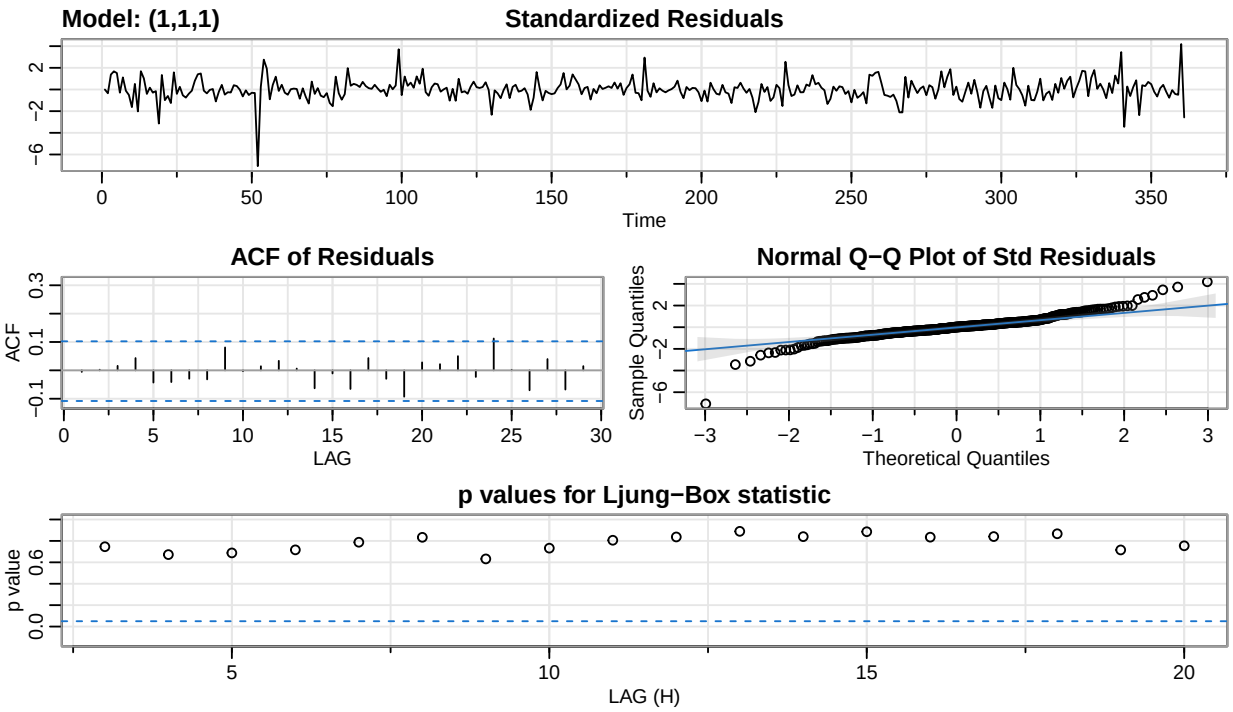

Figure 8: Model diagnostics for ARIMA(1,1,1)

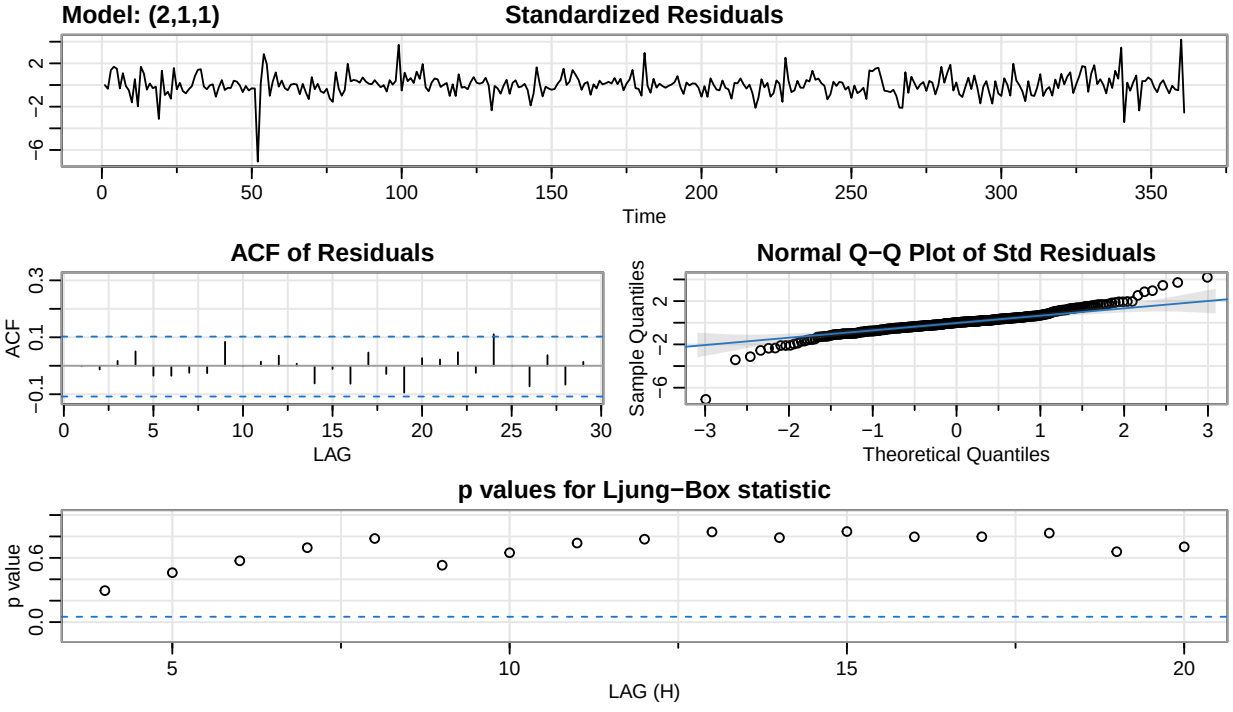

Figure 9: Model diagnostics for ARIMA(2,1,1)

3.3.2 AIC

Akaike information criterion (AIC) is a statistical tool to examine how a model fit the data, and the model's simpilicity, and it has been widely used to compare different model for selecting the best model[]. The lower AIC, the better the model fits the data. The AIC of ARIMA(1,1,1) is 110.95 and ARIMA(2,1,1) model is 112.81.

3.3.3 Parameters significance

To test the significance of the parameter we set a null hypothesis and a alternative hypothesis as follow:

- null hypothesis H_0 :the coefficients of the parameters are not significant different from 0
- alternative hypothesis H_A:the coefficients of the parameters are significant different from 0

Below tables shows the coefficients of each parameters and their p-value.

	Estimate	SE	t.value	p.value
ar1	0.4945	0.2760	1.7915	0.0741
ma1	-0.6375	0.2472	-2.5789	0.0103
constant	-0.0128	0.0106	-1.2073	0.2281

	Estimate	SE	t.value	p.value
ar1	0.5966	0.2971	2.0079	0.0454
ar2	0.0346	0.0889	0.3892	0.6974
ma1	-0.7453	0.2923	-2.5497	0.0112
constant	-0.0128	0.0102	-1.2565	0.2098

4 Result

4.1 ARIMA(1,1,1)

Figure 8 is the model diagnostic plots for the ARIMA(1,1,1) model, we observed that the standard residuals shows no obvious patterns, and there are no outliers. The ACF Residuals plot shows only one spike. Hence, we conclude that there are no violations of the model randomness assumption. In addition, there are no heavy tails on both ends of the Normal Q-Q Plot of Residuals, it supports that there are no apparent violations of the model normality assumption. Besides, all the p-values for Ljung-Box statistics are above the significant level, so there is no violation of the model independent assumption. So no model assumptions are violated. The ARIMA(1,1,1) model is a valid model with a Akaike information criterion (AIC) of 110.95.

Figure 9 is the model diagnostic plots for the ARIMA(1,1,1) model, we observed that the standard residuals shows no obvious patterns, and there are no outliers. The ACF Residuals plot shows only one spike. Hence, we conclude that there are no violations of the model randomness assumption. In addition, there are no heavy tails on both ends of the Normal Q-Q Plot of Residuals, it supports that there are no apparent violations of the model normality assumption. Besides, all the p-values for Ljung-Box statistics are above the significant level, so there is no violation of the model independent assumption. So no model assumptions are violated. The ARIMA(1,1,1) model is a valid model with a Akaike information criterion (AIC) of 112.81.

So both models did not violate any model assumptions. However, ARIMA(1,1,1) has a smaller AIC(110.95) than ARIMA(2,1,1) which has a AIC of 112.81. Thus from the AIC perspective, ARIMA(1,1,1) is the better model which may fit better to the data, and has more simplicity. In addition, the parameter tables suggest that parameter of ARIMA(1,1,1) has p-values of 0.0741 and 0.0103, thus we have no evidence to reject our null hypothesis, thus all the parameters in ARIMA(1,1,1) are significant. On the other hand, the

parameters of ARIMA(2,1,1) has p-values of 0.0454,0.6974, and 0.0112 respectively, while the first AR and MA parameters has small p-values, the second AR parameter has a p-value that is larger than 0.5, which means we can not reject the null hypothesis, so the second AR parameter in the ARIMA(2,1,1) model is not significant.

So in general, we found the ARIMA(1,1,1) is the better model to fit the time series.

4.2 Predicting the future price

Then we use the ARIMA(1,1,1) with $\phi = 0.4945$ and $\theta = -0.6375$ to predict the price of Dogecoin for the next ten days. Figure 10 visualize our prediction, however since we Box-Cox transformed our data in previously, the value of prediction has also been Box-Cox transformed.

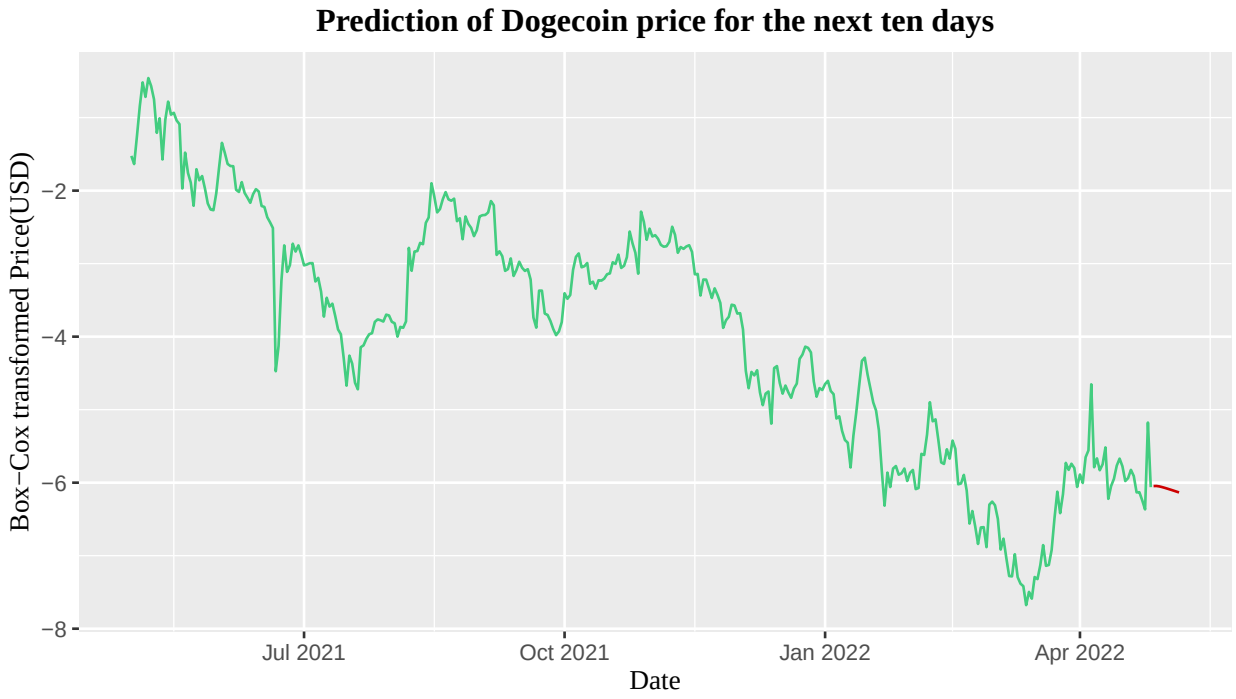

Figure 10: Dogecoin price prediction for the next ten days

We calculate the acutal predicted price which has not been Box-Cox transformed by doing a inverse Box-Cox transformation. Figure 11 shows the actual prediction of the Dogecoin price in the next ten days as from 26th April 2022. We observed that the price will be continuously decreasing for the next ten days.

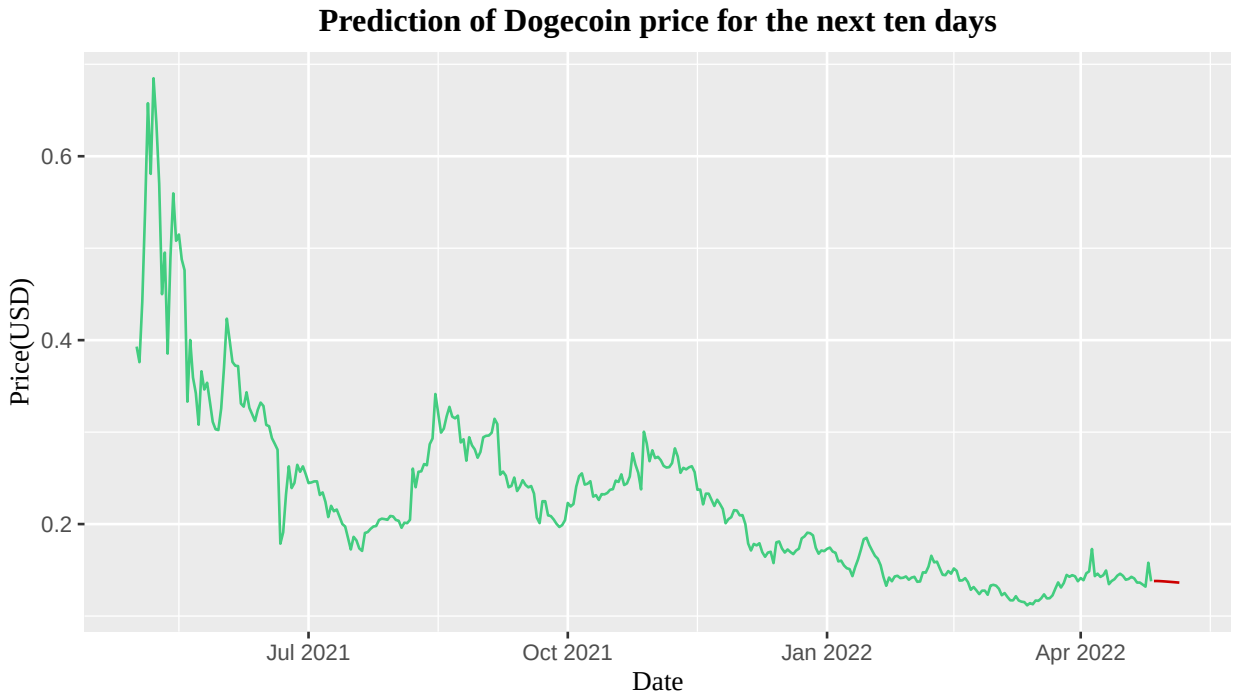

Figure 11: Dogecoin price prediction for the next ten days

5 Disscussion

5.1 Conclusion

We found that a ARIMA(1,1,1) model with coefficient $\phi = 0.4945$ and $\theta = -0.6375$ can fit the Dogecoin data from May 2021 to Apirl 2022, then if we apply this model topredict the future price of the Dogecoin we can find that the price will be coninously decreasing for the next ten days. So for investors who are looking for investment opptunuities, investing in Dogecoin may be a bad choice. In addition, the decreasing trend may also imply that Elon Musk is losing his interest in the Dogecoin and crptocurrency.

5.2 "All models are wrong"

A famous British statistican George E. P. Box once said "All models are wrong, but some are useful"[]. Since Dogecoin is highly volatile cryptocurrecy, its price could may be influenced by many things like government policy, financial influencer's involment, and so on, in some cases, a cryptocurrency may be soaring today, and disappear tomorrow. So our model certainly can not predict the actual future value of the Dogecoin, however the prediction can show investors some insights about the overall trend of the price to help them decide whether to invest it or not.

5.3 Next step

Except for the fact that cryptocurrency is naturally hard to predict, in our data, we only have useful 361 observations, in some cases, 361 observations is enough for a data to build a decent model, however, for financial prediction the more useful observattions the better. So for the next step, I will continue monitoring

the Dogecoin price to see if the actual price match our prediction, if not, I will try to build another model to fit the data.

5.4 Relevent research

I think antoher research which focuses on the building the relationship between the number of Elon Musk's tweets about Dogecoin and the price of the Dogecoin could be made, to see if there is actually a relationship between them, if so, how can we use this relationship to predict the Dogecoin price and make investment advice.