

Dogecoin to the ground

A forecasting of the Dogecoin price

Siming Shan

Contents

1 Introduction:	1
2 Data	2
3 Method	4
3.1 ARIMA model	4
3.2 ACF and PACF	4
3.3 Model diagnostics	4

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% [1]. 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 crypto currency that he bought - Dogecoin. Dogecoin is a open-source cryptocurrency created by Billy Markus and Jackson Palmer[2]. The cryptocurrency is a digital or virtual currency which is secured by cryptography[3]. 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 cryptocurrency 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 programming language R[4] and focus 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 dogecoin price in the next ten day. The data was collected from Yahoo finance by using the function `getSymbols` from the `quantmod` package[5]. In the original 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 unemployment 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 our 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

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 originally 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.

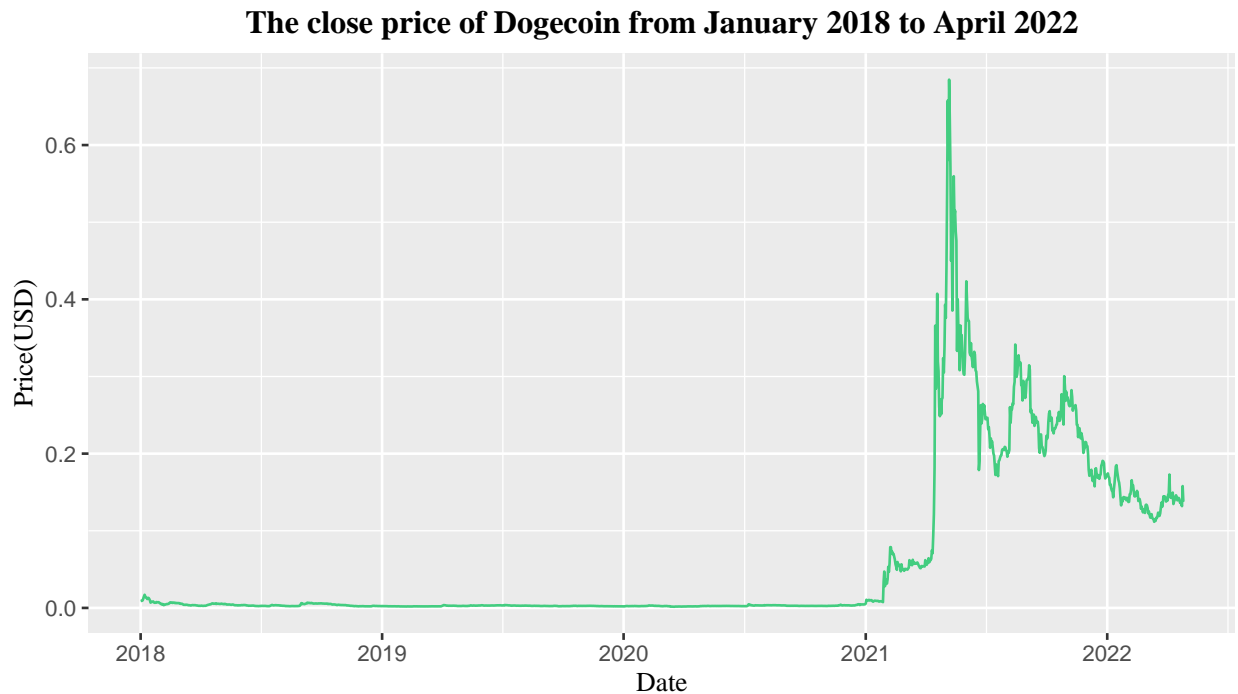


Figure 1: The close price of Dogecoin from 2018 to 2022

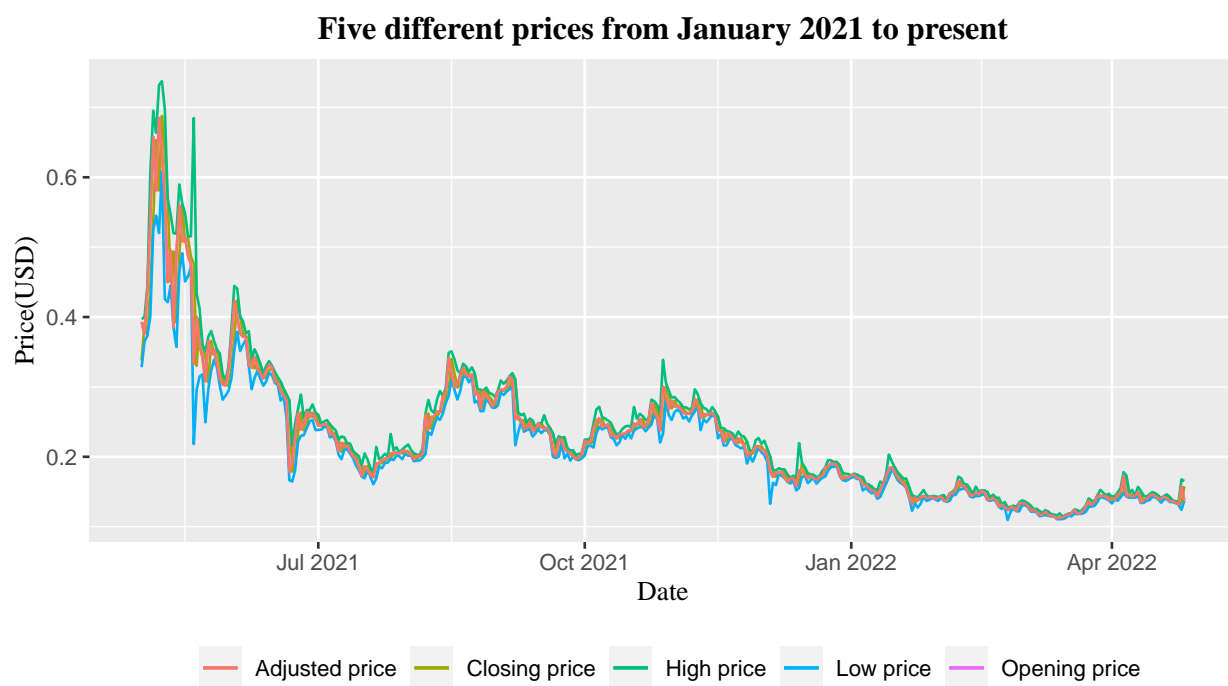


Figure 2: Five different prices of the Dogecoin from May 2021 to April 2022

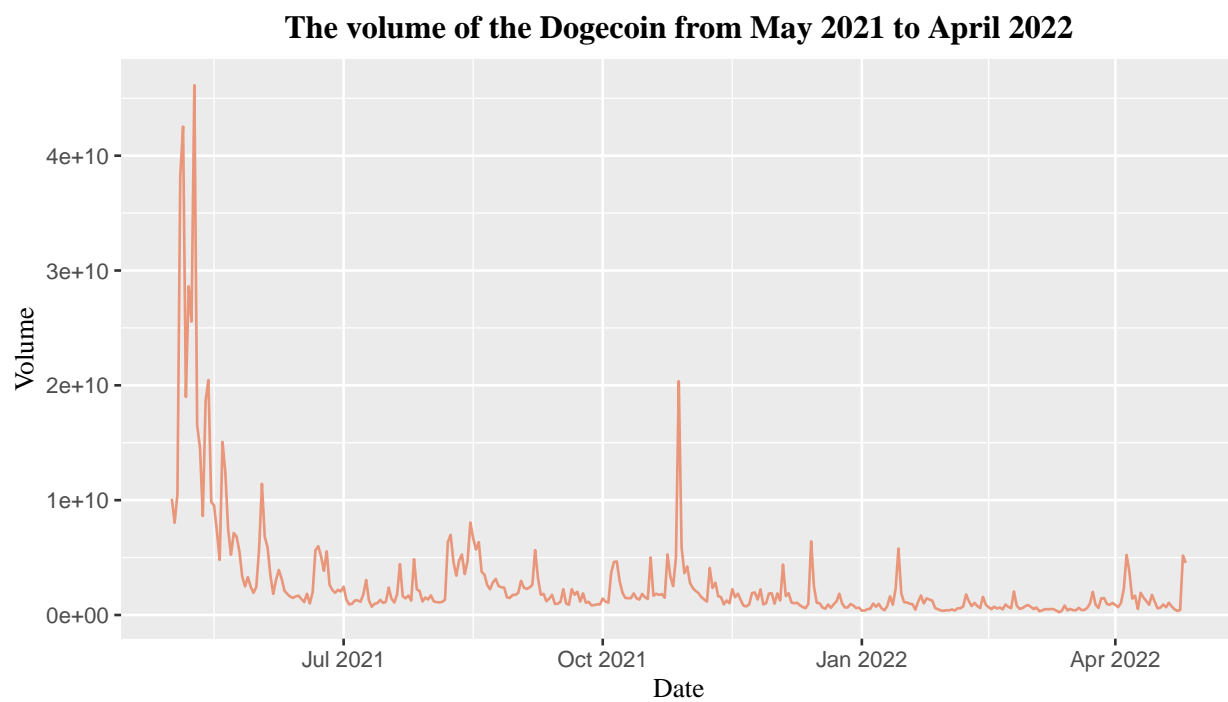


Figure 3: The volume of the Dogecoin from May 2021 to April 2022

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 is the moving average order. The mathematical equation of a ARIMA model can be expressed as follows:

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

In the above equation, X_t is the value to predict, $X_{t-1}, X_{t-2}, X_{t-3} \dots X_{t-p}$ are past values, ϕ_p are the coefficients of the past values, $W_{t-1}, W_{t-2}, W_{t-3} \dots W_{t-p}$ 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 seasonality 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 series is not stationary, thus a first difference transformation was applied on the data. The Figure ?? 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 stabilize the variance. The Figure 6 shows the price of the Dogecoin after both transformations, we observed that both mean and variance have been stabilized, 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.

3.2 ACF and PACF

An Autocorrelation function (ACF) plot and a Partial autocorrelation function (PACF) plot were used to determine the AR and MA order. Autocorrelation function calculates the correlation between one observation (X_t) with its lag (X_{t-h}), while PACF also measures 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 plot shows a gradually declining to zero after the first lag. The PACF also shows a gradual declining to zero after the first lag. So the AR order of the model could be 1 or 2 since PACF is also quite large at second lag, and the MA order of the model could be 1 since the ACF declining rapidly after lag 1. Thus two possible models can be introduced, ARIMA(1,1,1) and ARIMA(2,1,1)

3.3 Model diagnostics

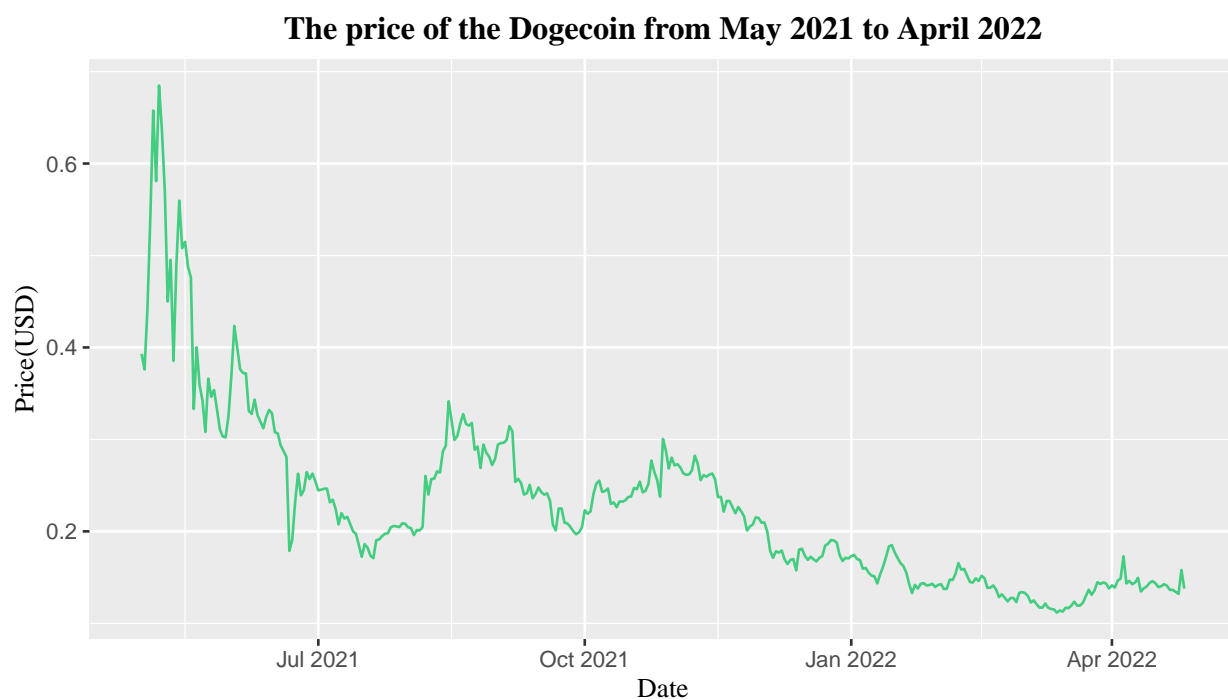


Figure 4: The price of the Dogecoin from May 2021 to April 2022

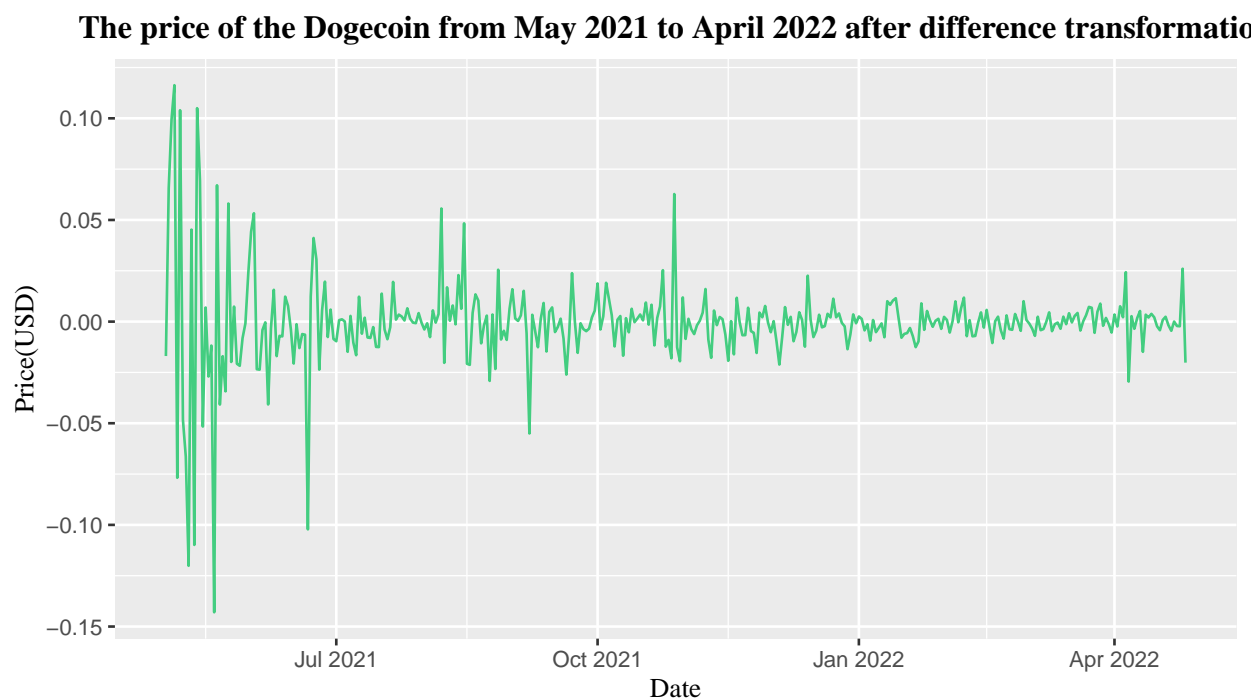


Figure 5: The price of the Dogecoin from May 2021 to April 2022 after difference transformation

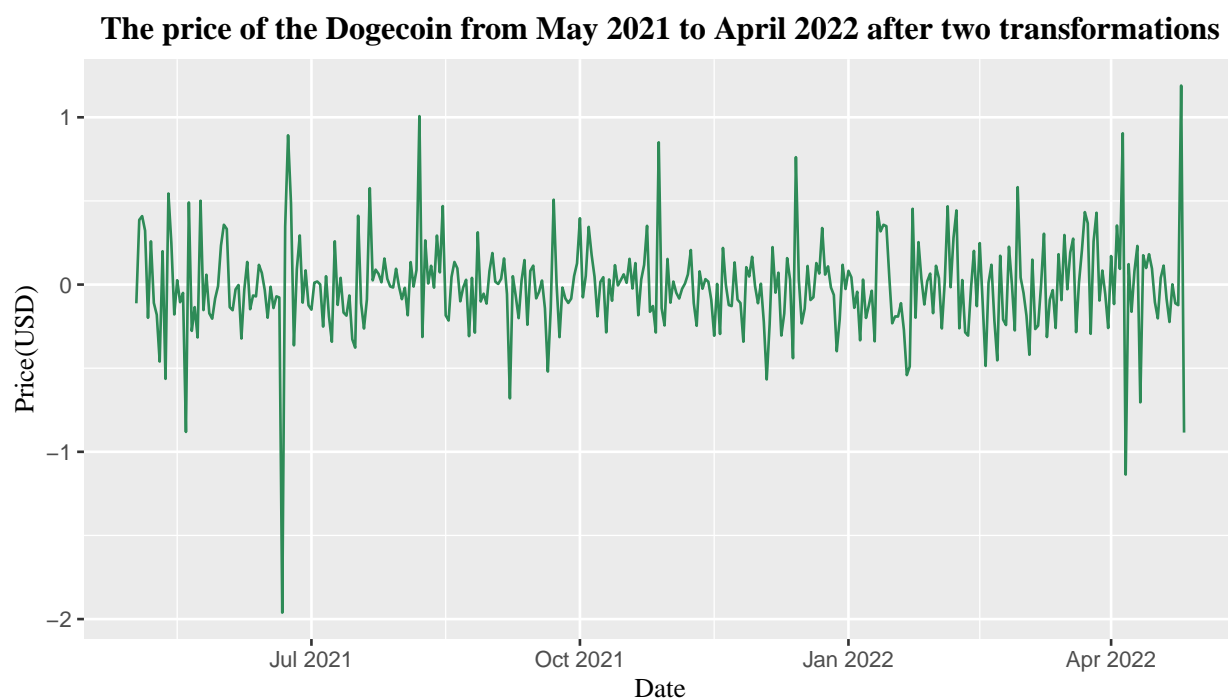


Figure 6: The price of the Dogecoin from May 2021 to April 2022 after difference and Box-Cox transformation

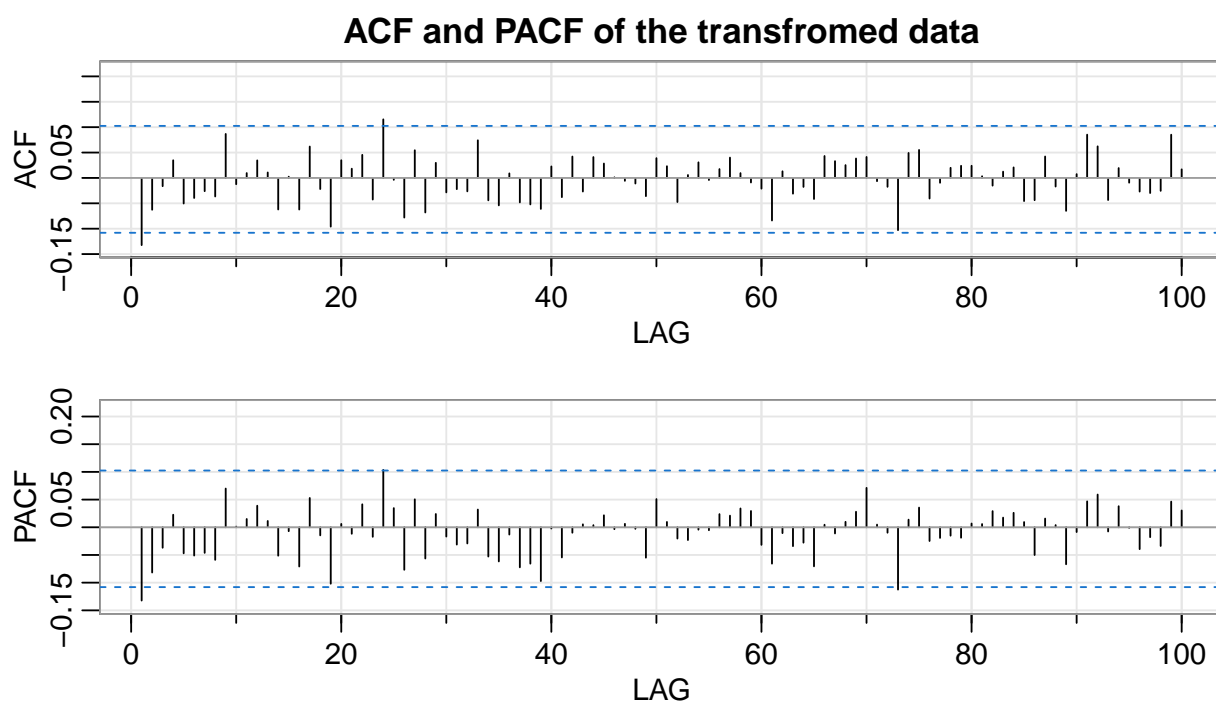


Figure 7: TACF and PACF of the transformed data