Data mining Bitcoin market for prediction

Zimin Lu  
computer science  
ST Thomas University *Miami,US*  
 zlu2@stu.edu

*Abstract*—In order to study the price trend of the bitcoin market and how to get the most benefit, a bitcoin price prediction was made.Use data from Yahoo Finance to build two models: arima and svr for prediction. The results show that the svr model has a high accuracy rate, can predict the price of Bitcoin in the future, and solves the problem of how to get most benefit in the Bitcoin market.

Keywords— bitcoin, price, prediction, model

# Introduction

This article is mainly talk about the data mining Bitcoin market for prediction. The purpose is to find out the method of predicting the price trend of the Bitcoin market through modeling of time series research, and obtain the maximum benefit by predicting the Bitcoins future price.

# Background

This part is related the data mining Bitcoin market for prediction’s background. It mainly includes the study of stock market previous studies, understanding the previous analysis methods of stock market prediction, exploring these methods are fit or not, and though their views develop my research methods.

## Related Works

Predict stock markets that mean forecast rate of return. For example, using the last year's return to predict the current stock return to determine which time you need to buy or sell.

In this article"Stock market index prediction using artificial neural network”[1], They used Nasdaq stock trading price data for the last 100 days, divided data by week then input these data. Use 70% of the data for training data, 30% of the data for testing data, training through ANN algorithm.Then they test two different test data(four prior days and nine prior days). Then choose the best network structure of the two data sets according to its prediction ability. And result R2(four prior days)=0.9408, the other one R2(nine prior days)=0.9622. They did not have a big difference. And using the ANN algorithm in this project can indeed get a high accuracy. In this case, they using deep learning methods can achieve high accuracy, so I have an idea of ​​the method used. Because stocks trend have the characteristics of time series, so that mean can use Support Vector Machines regression to make predictions.

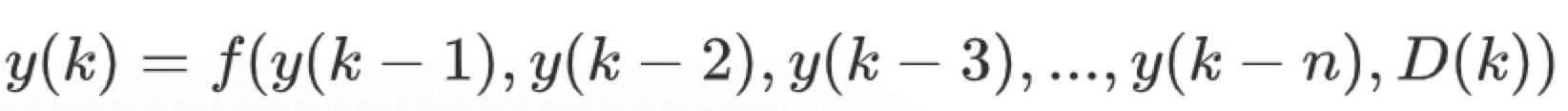
Another case comes from“STUDY OF EFFECTIVENESS OF TIME SERIES MODELING (ARIMA) IN FORECASTING STOCK PRICES”.[4]Their data includes 56 different industry stocks from the Indian stock markets. Then using Akaike information criterion(AIC) to comparison and parameterize the ARIMA model. When comparing more models the one of AIC’ s value get the lowest, then that model is more closely with real data. In this case,They used the ARIMA model related to time series, which also confirmed my conjecture that the ARIMA model can indeed predict stocks.

## Related Approach

ANN

For predict stock markets return method the researcher more like to use deep learning to study that. And the ANN(artificial neural network) algorithm is the most common one. Because ANN is based on the brain's processing mechanism, it is suitable for developing algorithms for building complex models and predicting problems. So the stock return prediction problem is very applicable to the algorithm.

On the ANN predict stock market’s article, researchers use feed-forward artiﬁcial neural networks that were trained by the backpropagation algorithm. About the variables they set up into this “where y(k) is the stock price at time k, n is the number of historical days, and D(k) is the day of the week”.[2] And the formula is:



For the build model, they used MATLAB software R2010a to determine the coefficient (R2) and the mean square error (MSE) of the model output to evaluate the performance of the ANN.

ARIMA

About predicted stock market time series also a good approach to research. Because stock return data is usually recorded by times. And ARIMA models is the most general class of models for predict a time series. Shows as “A random variable that is a time series is stationary if its statistical properties are all constant over time.”[3] That mean when observing stock changes, you can clearly see the changes in price trends.

Initially, an appropriate ARIMA model must be identified for a specific data set.And the parameter should have the smallest possible value so that the data can be analyzed correctly make predictions accordingly.

The two different algorithms have their own advantages. The ANN algorithm has random sampling and obtains the best model through testing. The ARIMA model is based on chronological order to find the regular pattern of change, and predicts future changes. Therefore, these two algorithms have high advantages in the research of stock prediction.

HMM

Hidden Markov Model (HMM) is a special case of the state-space model, where latent variables are discrete and polynomial. And the ability is to predict and analyze time-based phenomena, so very useful in areas such as financial market forecasting. This another case is about how to build a stock prediction model with HMM“Stock Price Prediction Using Hidden Markov Model”.[5] They get three years of historical stock price data using Pystock data and push data to GitHub repository, and data gather by day. Create a CSV file containing all historical data for GOOGL, FB, and AAPL stocks, Create a prediction model called StockPredictor, using the change in scores of each stock used to train the HMM.Use the hmm model to train and calculate Xt + 1 = <fracchange, frachigh, fraclow> Observe the vector, which is to find Xt + 1 that maximizes the posterior probability of P (Xt + 1 | X1,…, Xt, θ) the value. Then draw a price curve.

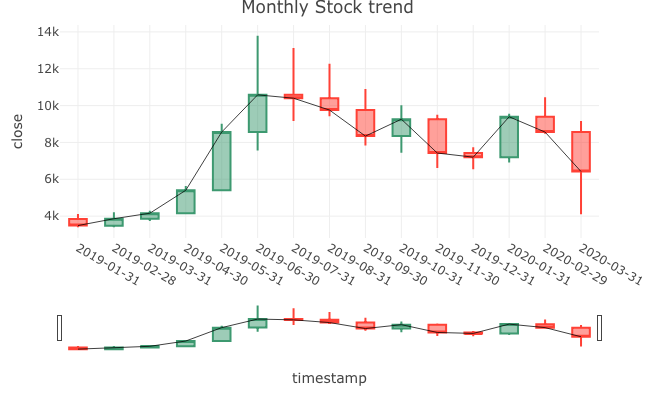
## Develop

## The research about “Data mining Bitcoin market for prediction”also will using deep learning prediction. The data set is derived from Yahoo Finance, including the opening price and closing price of a day and the highest and lowest values ​​over a while.So we can use the LSTM network for time series prediction based on Keras, Support Vector Machines regression and also can use ARIMA. Use the closing price and volume as the x input and the opening price of current the day as the y output.

# Exploratory analysis

Before building the model, I will visually analyze the data and summarize the key variables to be analyzed.

In this part will show a few pictures to description changes in the BTC market.And data from yahoo finance.

1.This candle chart shows monthly of Bitcoin transaction trend. Through the trend the Bitcoin buy and sell between 2019-04-30 to 2019-06-30 showing a sharp upward trend,then there has a downward trend. As of 2020-03-30, the close price has dramatically drop.Judging from the trend, it may continue to fall on next month.

2.On the weekly candle chart the mainly change start on 2019-04-07, this is first time Bitcoin close price get to 5000 and rose all the way. Since 2019 April,Bitcoin at a stage of good development time.



3.In spite of the bitcoin market show the down trend on 2019 mid-July but there still has nice price.

However on daily trend plot,there has a significant decline on 2020-03-12.On this day the open price is 7913.616 but the close price is 4970.788.Drop nearly 3000 points.After that the price never more than 7000 point.This may mean the decline of the bitcoin market.



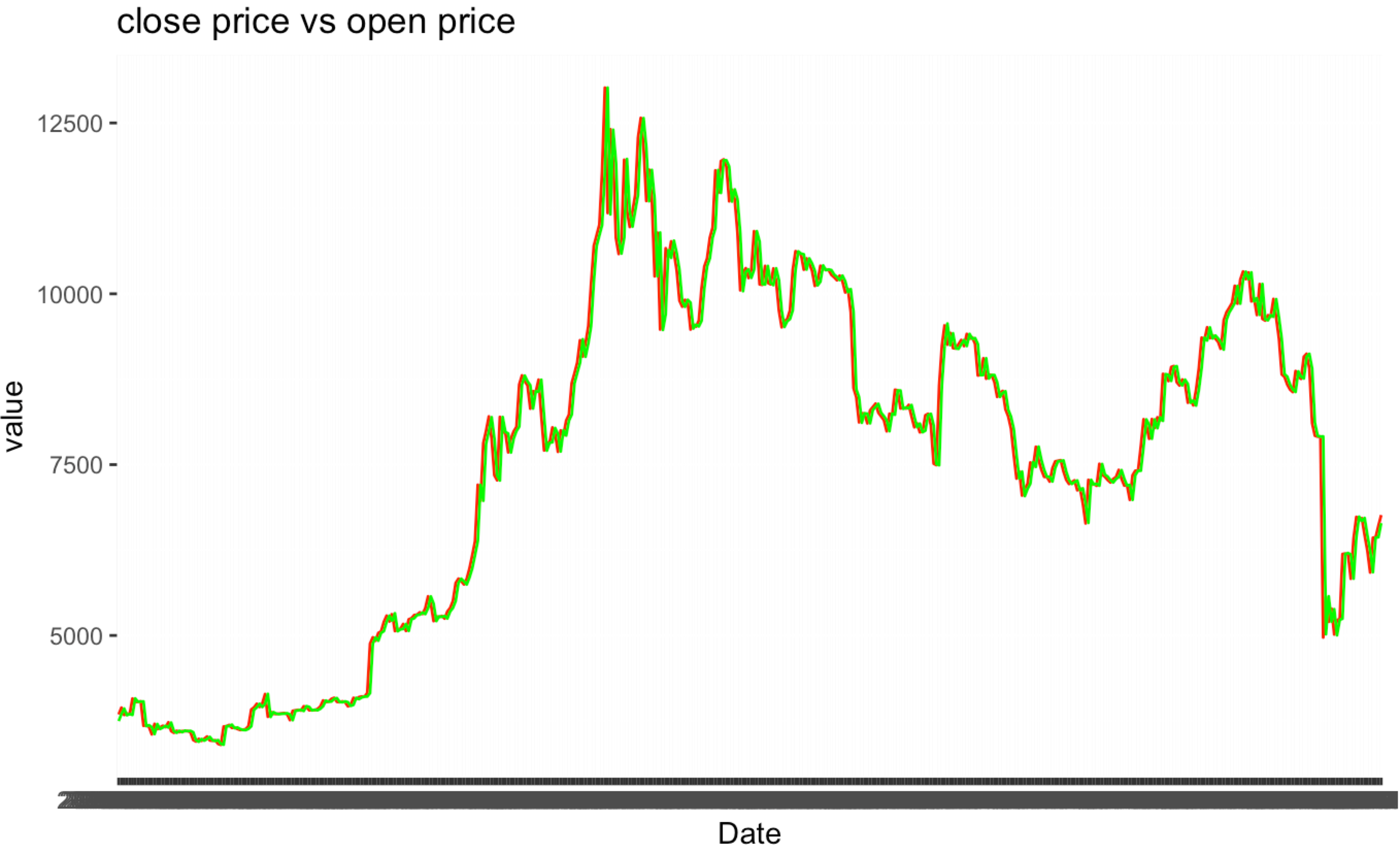
4.In this plot, most point gather on the left side, that mean with rising volume, stock prices are on the rise, there has a positive correlation between trading volume and stock price.

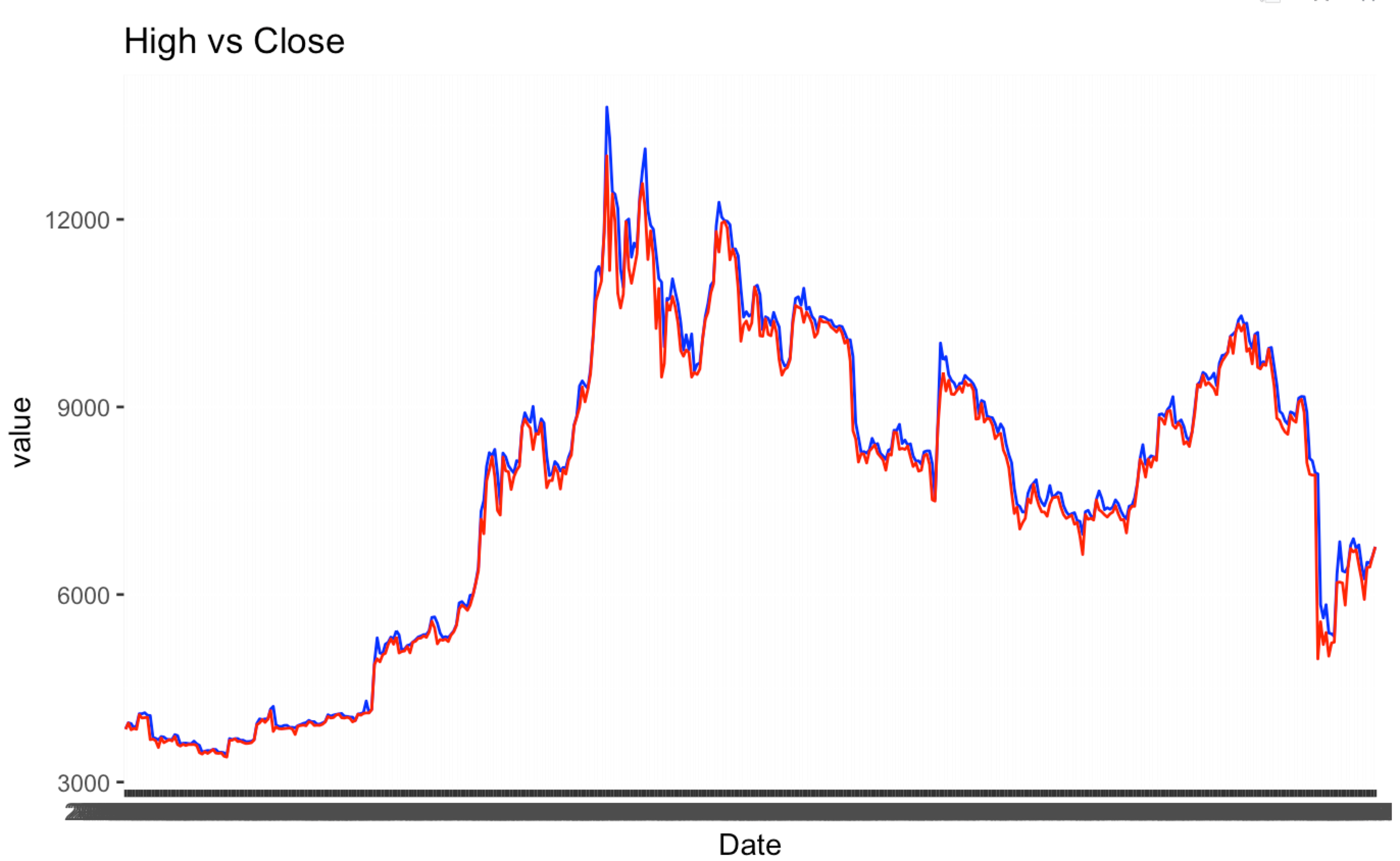
## 

5.This plot shows the relationship between high price and close price.When the price high people buy the stock because that time they can get more benefit, so the high price line(blue) looks complete same with close price line(red).

And displaying the changes in data for one year is conducive to a more intuitive understanding of the changing trend of the Bitcoin market from low to high prices.

6.In stock market the close price is the last transaction price on the trading day.And open price is the first transaction price of a security after the market opens on each trading day.If a certain kind of securities is not bought or sold within half an hour of market opening, the closing price of the previous day is taken as the opening price of the securities on that day. So this chart can predict future stock market conditions based on price fluctuation curves.Basically the closing price is the same as the opening price.

So using the closing price prediction can get the opening price of the new day.

7.In this part I make a formula to get bitcoin price and Price fluctuation. In this plot there has two strong change part, correspond "high price vs close price trend" this plot two part of rise and plunge.In general, the bitcoin market is going through a variable period.

# 

# IV. Methodology and model design

## Data sources

The preliminary data was obtained from the Yahoo Finance. Historical records about price number of bitcoin transaction at this website, data that goes back to 2019 until 2020. And can query the data for each day.These structured data provide detailed information for modeling.

## 5Vs of Dataset

Volume: Monthly data is from 2019 January to 2020 April per month.

Weekly data is from 2019 March to 2020 April per week.

Daily data is record 2019 to 2020 per day.

Variety: All data comes from Yahoo Finance and all data is structured data.

Value: Data records stock market volatility, these data has great help in forecasting trends.

Velocity: These data records the current day.

Veracity: These data come from authoritative financial websites, so they have high quality, high credibility and value.

## C. Variable Description

### *Independent Variables*

Close (numeric)

Description: Bitcoin market close price

High (numeric)

Description: Bitcoin market high price

Low (numeric)

Description: Bitcoin market low price

Volume (numeric)

Description: Bitcoin market buy and sell quantity

Price (numeric)

Description: Bitcoin price

### *Sub-Independent Variables*

Date (numeric)

Description: Bitcoin market trading hours

### *Dependent Variables*

Open (numeric)

Description: Bitcoin market open price

Adjusted (numeric)

Description: Bitcoin market adjusted price, adjust according to closing price

Diff-price (numeric)

Description: Bitcoin price fluctuation

## D. Algorithm parameters

### *Algorithm choose*

ARIMA

SVR ：SVR is same as SVM, can use regression analysis to predict the target based on the values ​​of these training samples.

These algorithms can handle timing issues, and bitcoin market data record by day, so these two algorithms are suitable for the research of Bitcoin market trends.

### *Parameters choose*

Date

Model building based on time.

Close price (numeric)

Open price (numeric)

The closing price affects the opening price, which can be predicted from the closing price.

High price (numeric)

Low price (numeric)

These two values ​​determine the price fluctuation.

Price

Bitcoin price

### *Training set*

Use one year data to being training set.

Use high price subtract low price, get the value divide by open price, and that value is price fluctuation.

When the known closing price can predict the opening price of the next day. When the price fluctuation show the positive value more high the price increase. If show the negative value the price decline.

Validation set

Use 2020 March data to being validation set.

Use these part data to test the model for 93 times and choose the best model, this model need has Minimal difference from actual data.(SVR part)

Test data:

Use 2020 April data to bring Testing set.

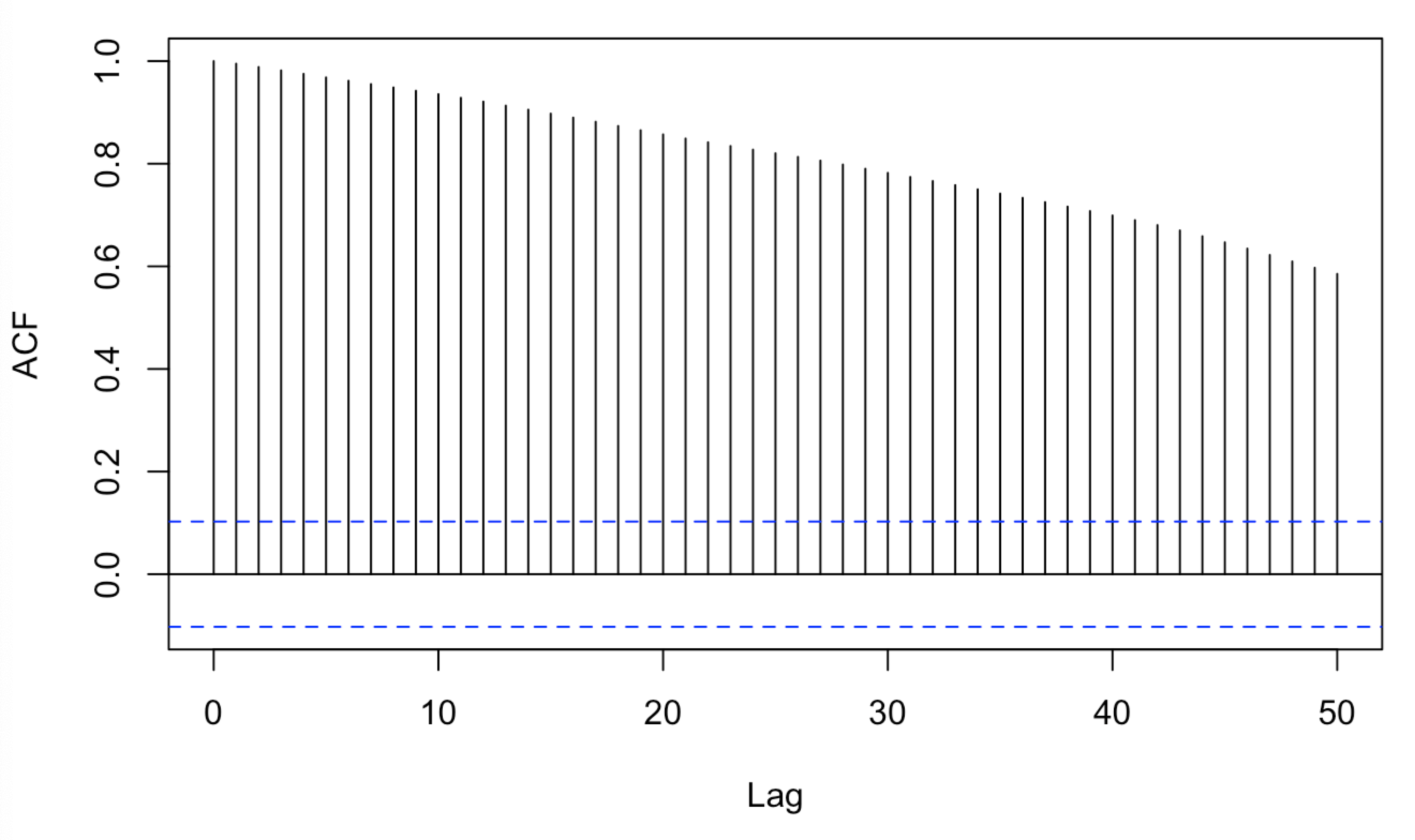
Use these part data to test model for 26 times, and get the point of price in 2020.4.28.

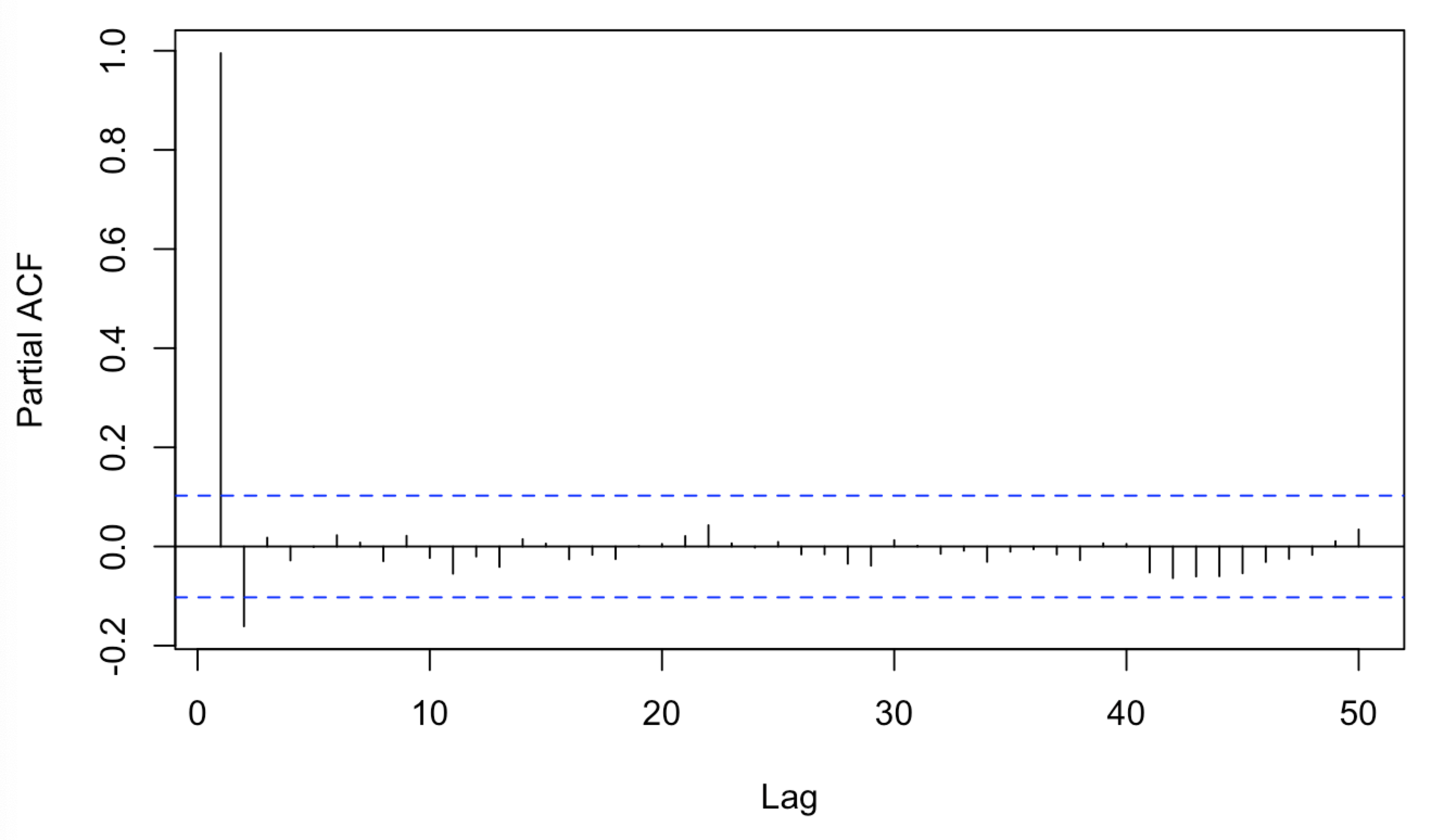
# V. Model building

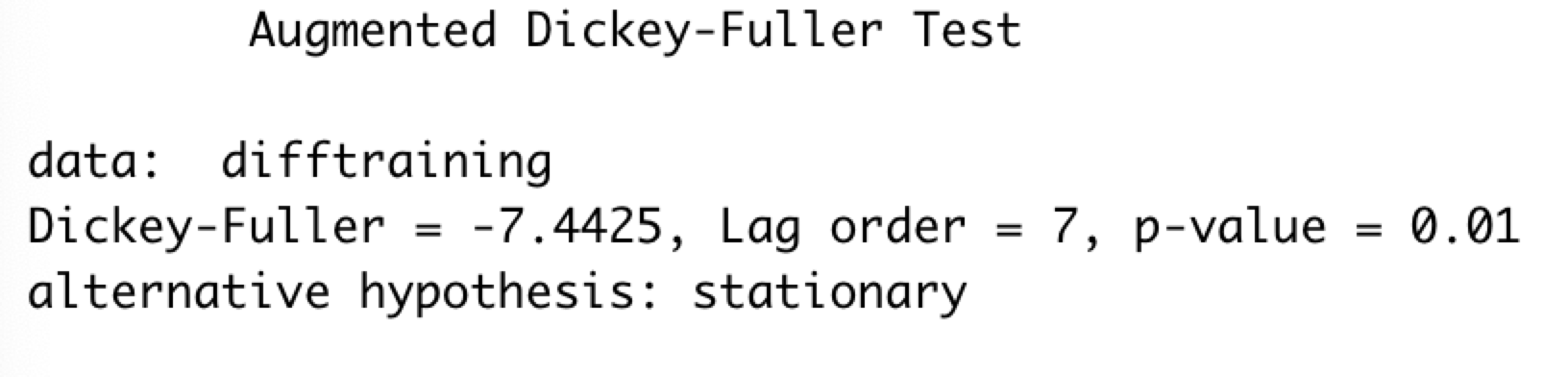
# A. arima

Because stock data is recorded based on time, I first adopted the ARIMA model.The entire data set contains 458 rows data. I will use the first 365 rows of raw data as training data and 93 rows of data as testing data.

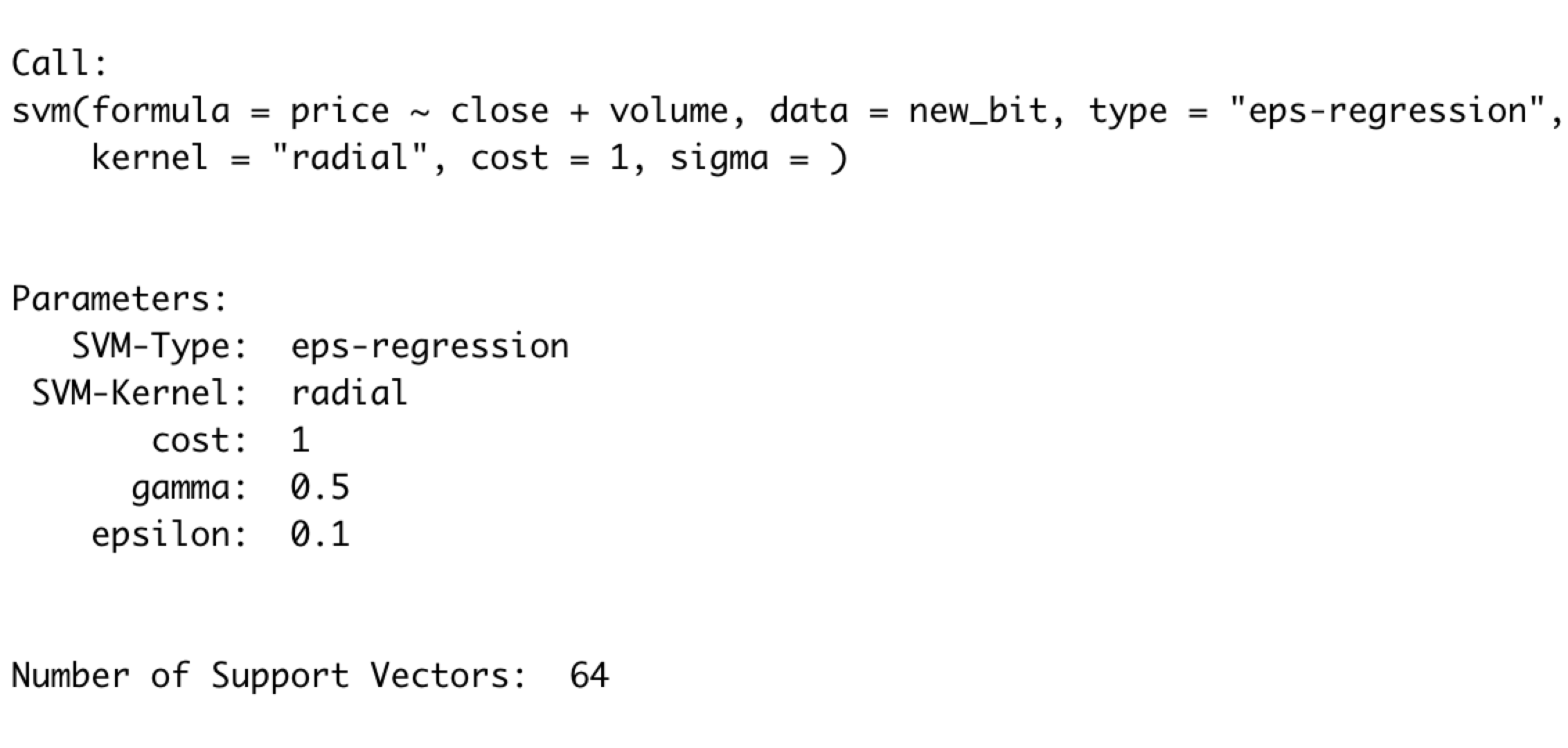
First check acf and pacf to determine q and p, then make a plot.

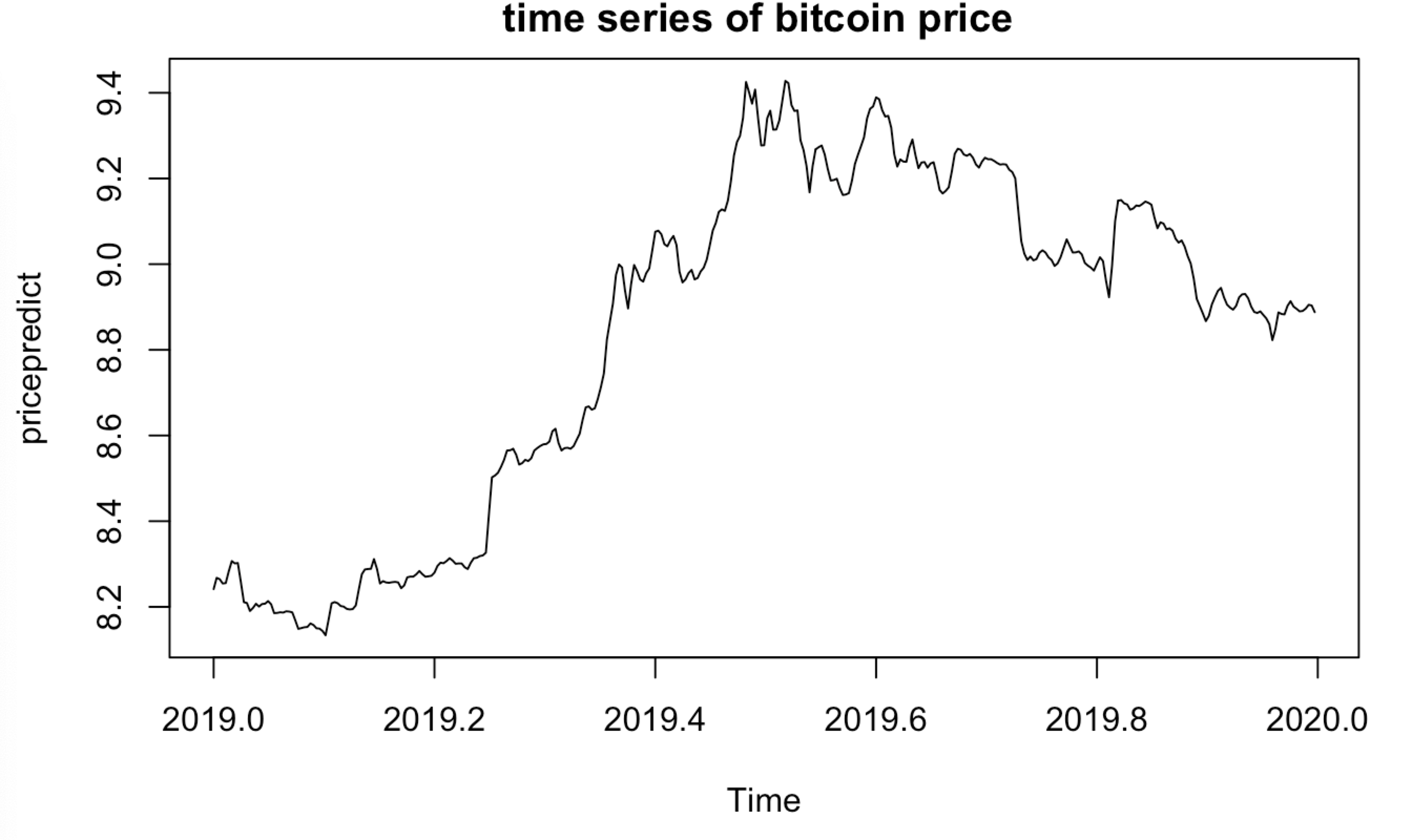
ACF plot:

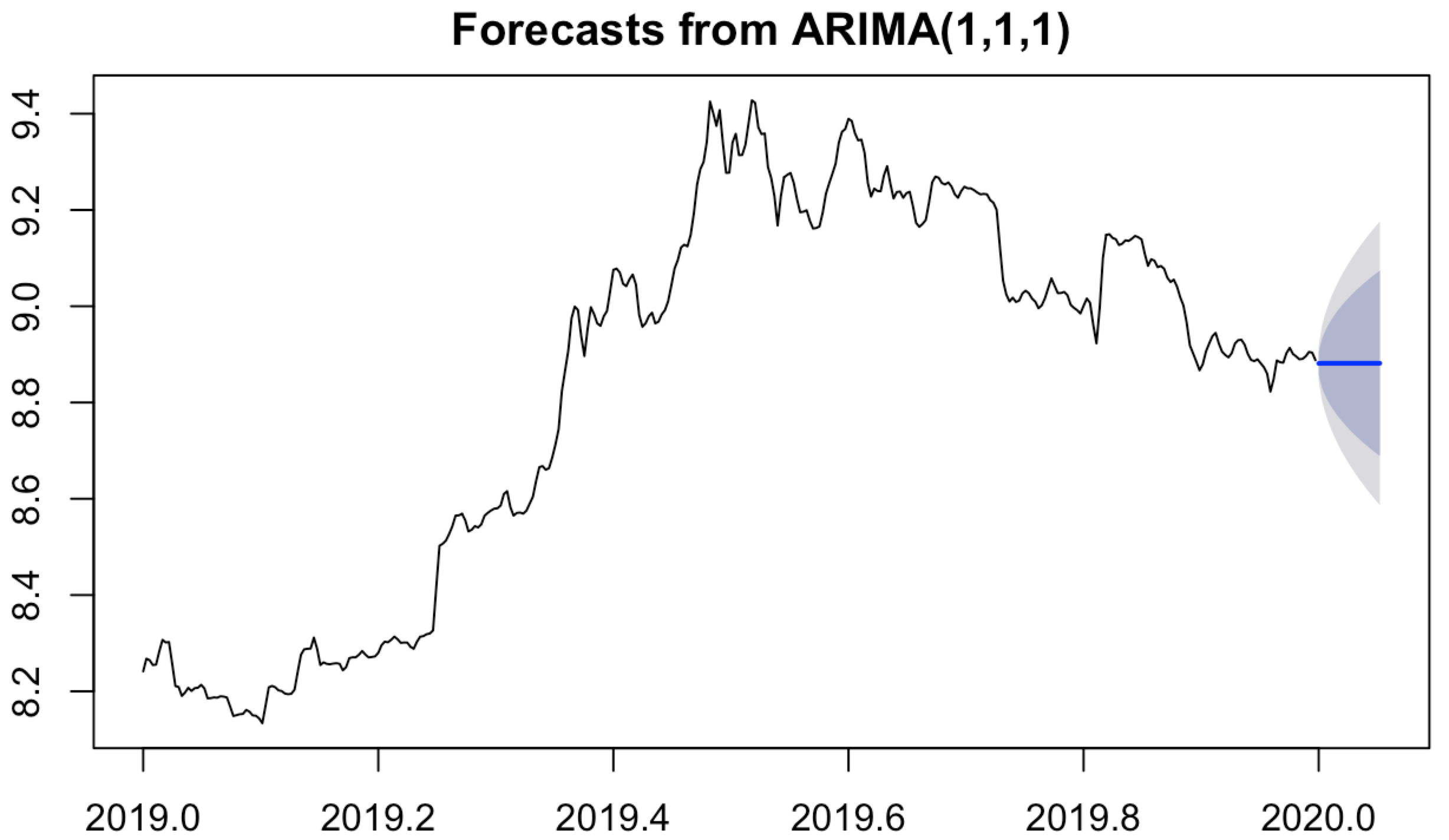
PACF plot:

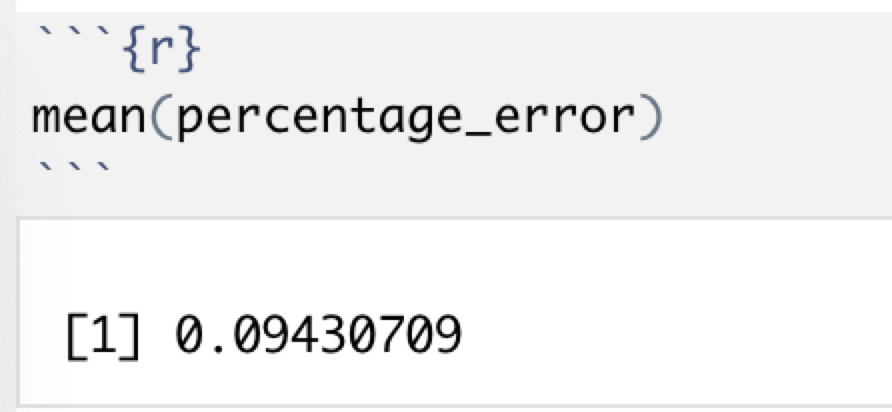
When the value=1 there has a peak, so the diff=1, next use ADF test to check diff-training data stationarity, and p-value=0.01, shows the series is static.

Identify applicable funding agency here. If none, delete this text box.

Use ARIMA make prediction and return the coefficient.Then make a plot.

Plot of Time series:

Make forecasted values from ARIMA:

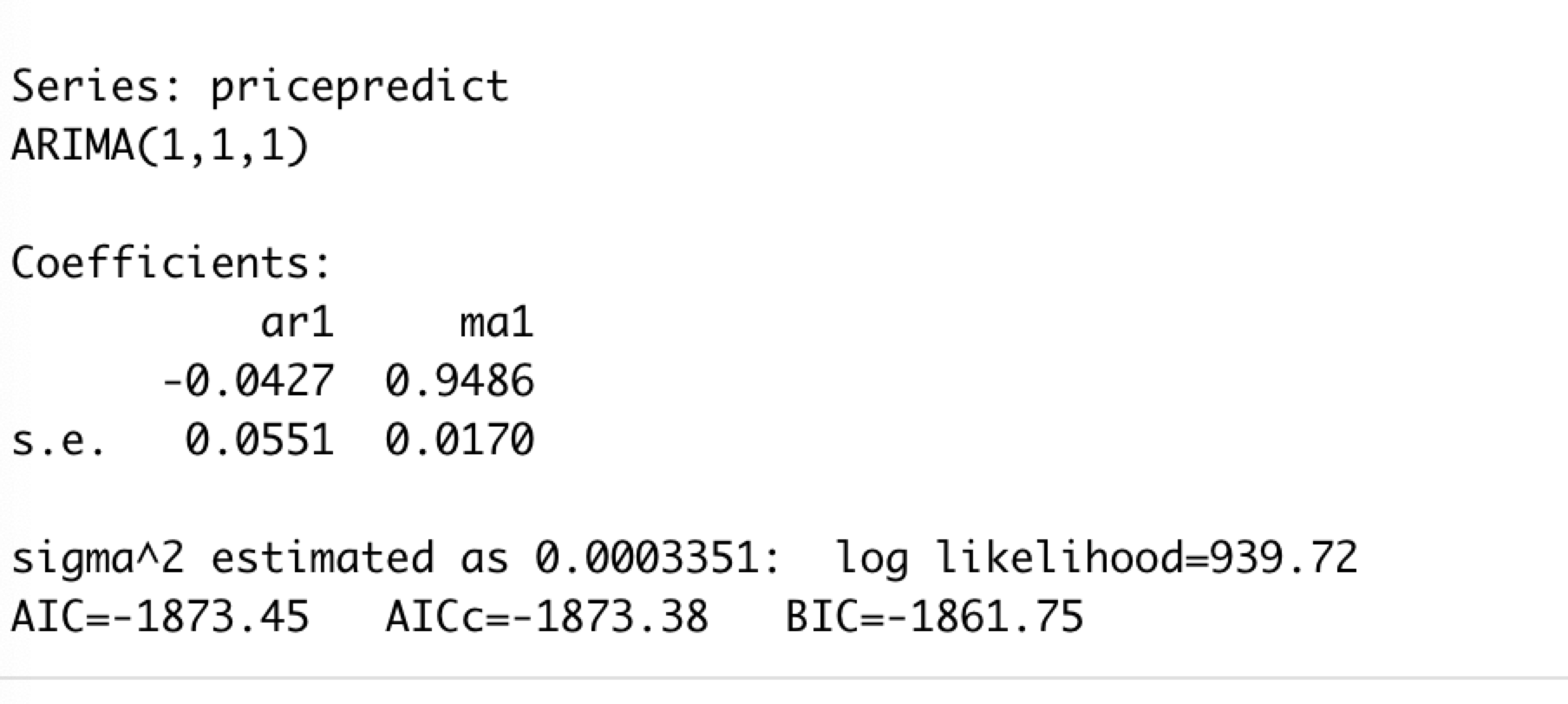
In the end make a percentage error to check the model is work or not.

The error shows 9%, so the model is work.

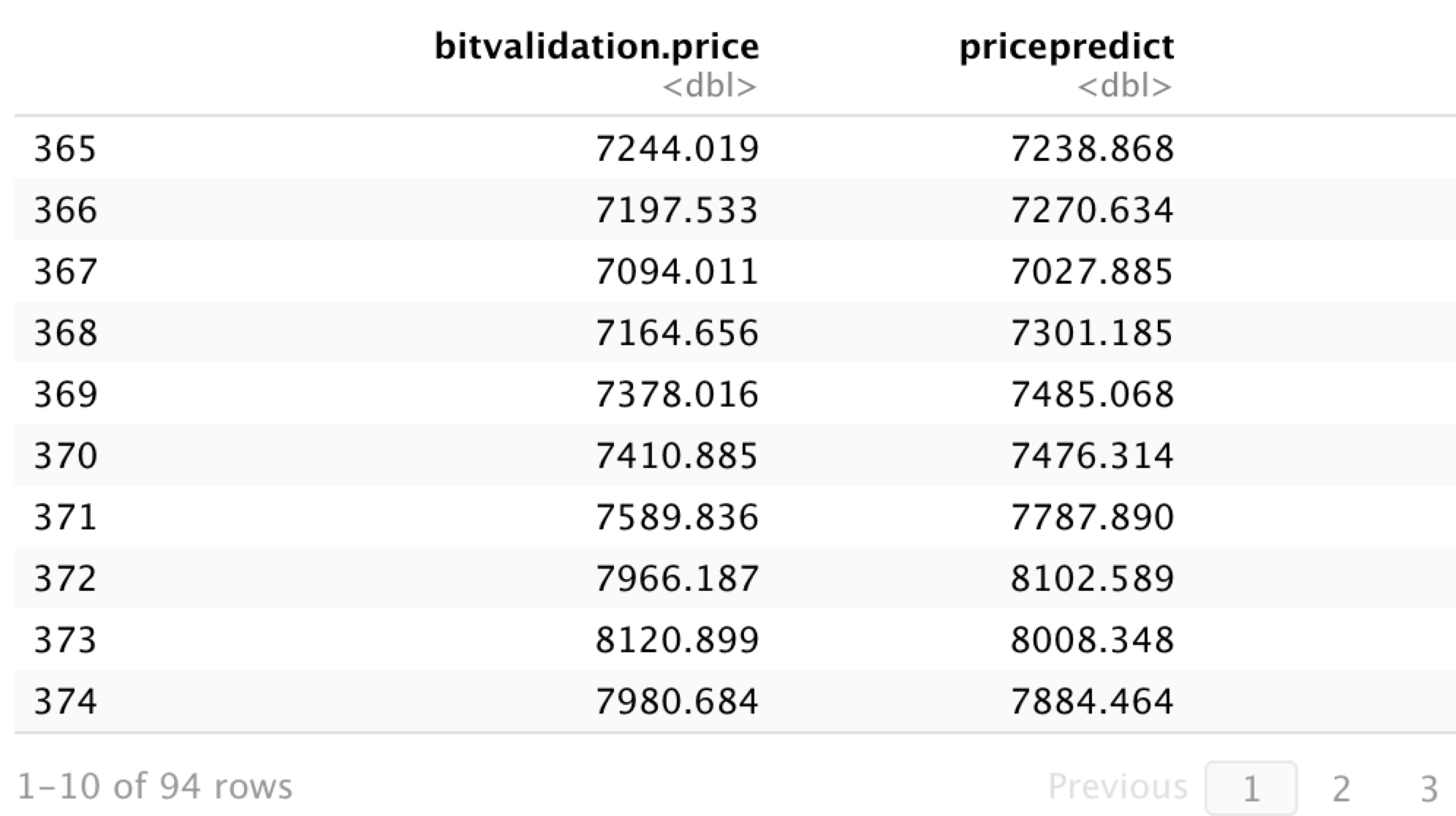
## B.Support Vector Machines regression

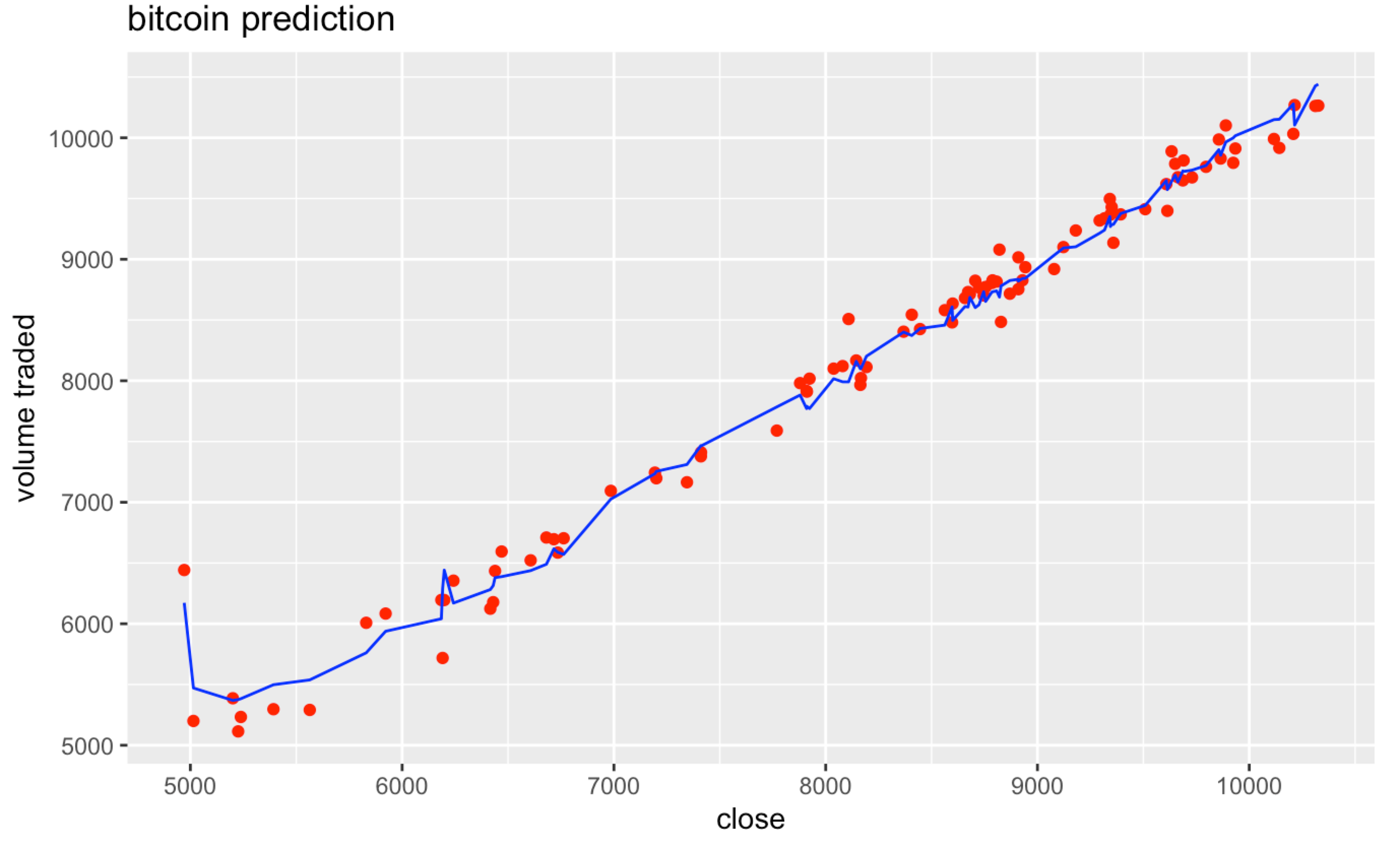
About using SVR build model, first need to divided the training set and testing set.

In this model I also use 2019 data into training data, and 2020 data into testing data.

Because the closing price determines tomorrow ’s opening price, and volume is a statistic of the amount purchased in a day, so can use the closing price and volume to predict tomorrow ’s stock price. And use svr() function can get the regression like that :

Then using regression result and testing set to make a prediction. And make a data form to compare the actual price between forecasted price.(part of form) When the closing price on day 365 is 7244.019, the predicted opening price on day 366 is 7234.968, while the actual opening price on day 366 is 7197.533.

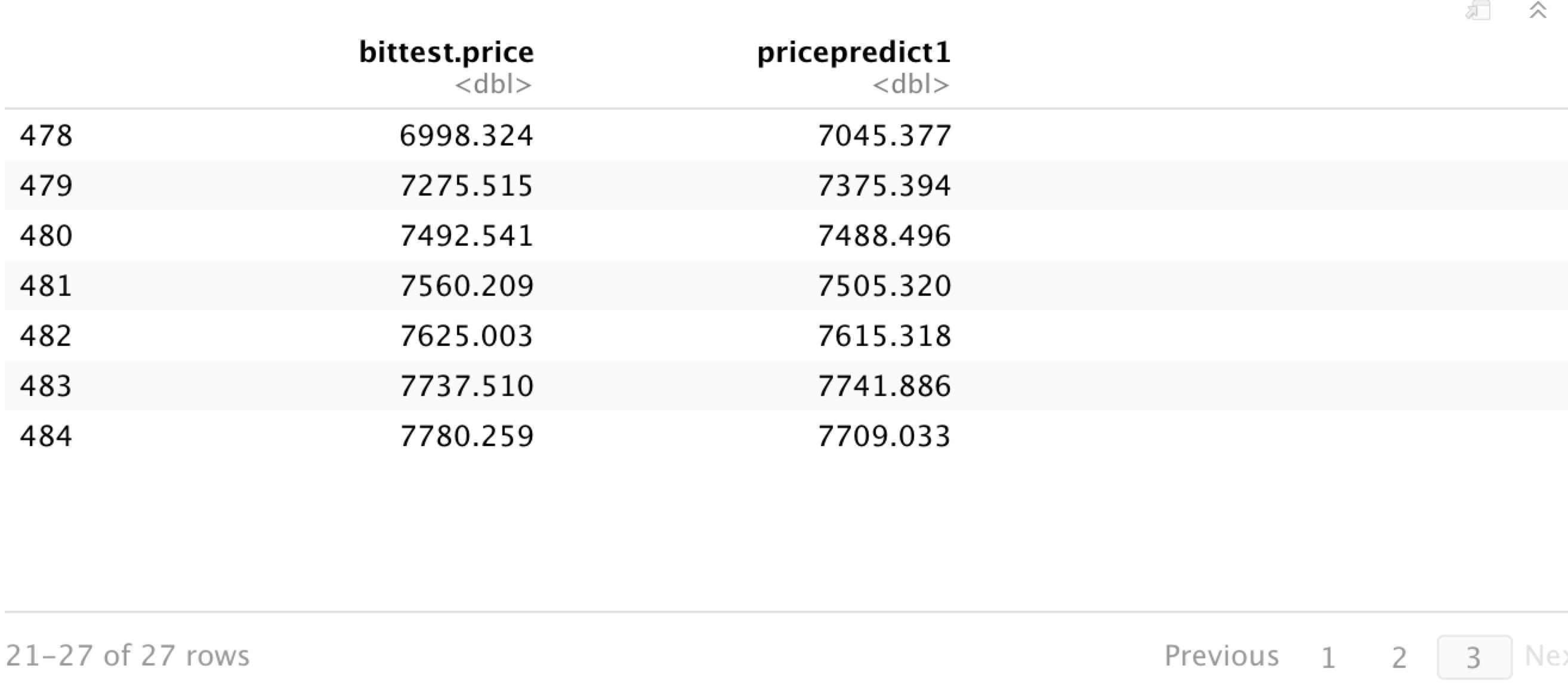


Prediction vs actual price shows on ggplot:

In this plot the actual point is close to predict line. So this model is also work and can knew the specific price. Now can use this model with testing set.

In this data form No.484 row is 2020.04.27, and This line predicts the day of 2020.04.28. From the prediction results 2020.04.27 the close price is 7780.259, 7709.033 is the opening price of 2020.04.28.

Judging from the results, this is a downward trend, which is suitable for selling transactions on 2020.04.28.



# VI. conclusion

Regarding the establishment of stock prediction models, both ARIMA and SVR can work. The advantage of ARIMA is that there can be a 9% error, but SVR can predict future specific data and the error is also small.

In the study of the future price prediction of Bitcoin, SVR model has a better effect and can solve this problem.

## References

Identify applicable funding agency here. If none, delete this text box.

1. A. Moghaddam, M. Moghaddam, and M. Esfandyari, "Stock market index prediction using artificial neural network", 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2077188616300245. [Accessed: 23- Mar- 2020].
2. A. Moghaddam, M. Moghaddam and M. Esfandyari, "Stock market index prediction using artificial neural network”.
3. R. Nau, "Introduction to ARIMA models", *People.duke.edu*, 2019. [Online]. Available: https://people.duke.edu/~rnau/411arim.htm. [Accessed: 23- Mar- 2020].
4. P. Mondal and S. Goswami, STUDY OF EFFECTIVENESS OF TIME SERIES MODELING (ARIMA) IN FORECASTING STOCK PRICES. pdf, 2014.
5. "Stock Price Prediction Using Hidden Markov Model | Rubik's Code", *Rubik's Code*, 2018. [Online]. Available: https://rubikscode.net/2018/10/29/stock-price-prediction-using-hidden-markov-model/. [Accessed: 24- Mar- 2020].