NYSE Stock Prices - Using Altaba's Stock Price Data As an Example

Yiqu Ding

12/3/2021

Abstract

This report analyzes data of earnings of Altaba's stock performance from 2006 to 2018. Using auto fitted auto regression moving avegrae integrated models, this report uses a SARIMA(2,1,2) to forecast Altaba's future performance. The model predicts a downward trend for the first differentiation of the stock prices.

Keywords: ARIMA, Time Series, Economics, Stocks

Introduction

Nowadays, the stock market is an essential measure of a country's economic status; therefore, analyzing stock returns is critical to understanding the economic trends. Stock price data is a typical type of time series because they are prices of a particular stock across a period. In this report, we fit four autoregression integrated moving average models on the closing price of stock AABA and generated four forecasts based on four autoregression.

Table 1: Data Summary

Date	Open	High	Low	Close	Volume	Name
Min. :2006-01-03	Min. : 9.10	Min. : 9.48	Min. : 8.94	Min. : 8.95	Min. : 1939061	Length:3019
1st Qu.:2009-01-01	1st Qu.:16.18	1st Qu.:16.39	1st Qu.:15.97	1st Qu.:16.13	1st Qu.: 12480251	Class :character

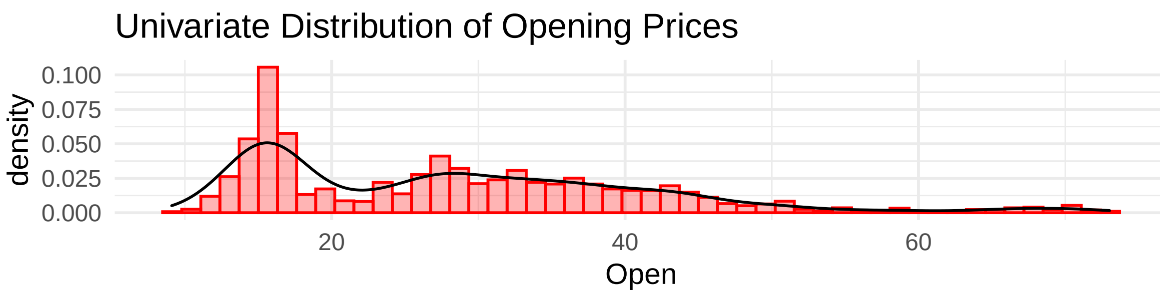
Date	Open	High	Low	Close	Volume	Name
Median :2011-12-30	Median :27.18	Median :27.49	Median :26.82	Median :27.10	Median : 17321301	Mode :character
Mean :2012-01-01	Mean :28.43	Mean :28.77	Mean :28.07	Mean :28.41	Mean : 21583912	NA
3rd Qu.:2015-01-01	3rd Qu.:36.66	3rd Qu.:37.03	3rd Qu.:36.30	3rd Qu.:36.63	3rd Qu.: 25127569	NA
Max. :2017-12-29	Max. :73.02	Max. :73.25	Max. :72.46	Max. :72.93	Max. :438231658	NA

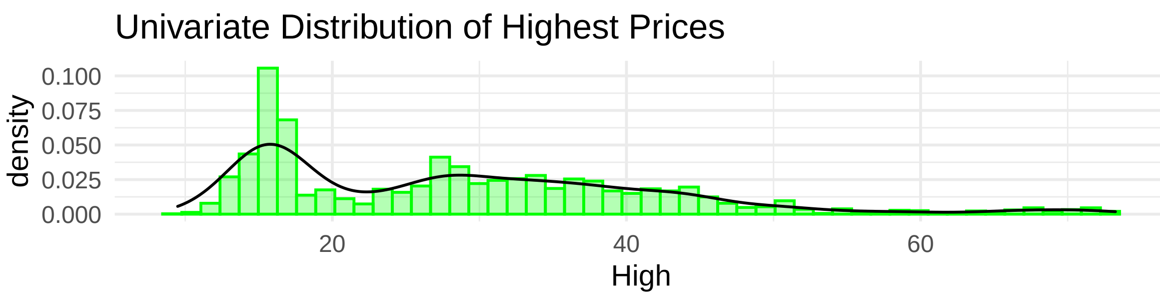
Data

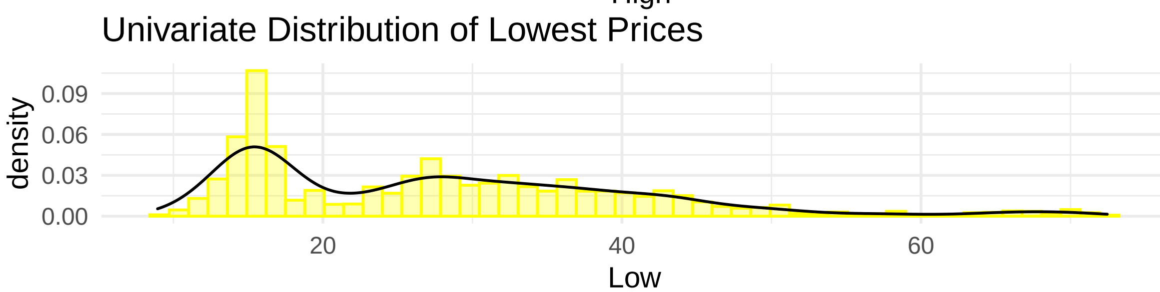
In this report, we use publically available data obtained from Kaggle.com to demonstrate the analysis of stock prices of Altaba. Inc, an investment company, based in New York City. The stock prices are recorded every day from 2008 to 2016 at the time point of market opening, market closing, and on every day's highest and lowest. There are 3019 observations with 7 variables:

- Date: an R date variable containing the date;
- Open: the opening price of a particular date;
- Close: the closing price of a particular date;
- Highest: the highest recorded stock price during the trading window of each day;
- Lowest: the lowest recorded stock price during the trading window of each day;
- Volume: the number of stocks;
- Name: a ticker for the stock;

Below are some exploratory data analyses:







Observe that the four series: Open, Close, High and Low, share similar univariate distributions, with a dominant peak around 15 dollars and a secondary peak between 25 to 30 dollars. Generally, the distribution is right-skewed. We show the time series analysis using closing prices, but the method should apply to all prices.

Method

We check that the model is stationary. With `adf.test()` which performs augmented dickey-fuller test, we see that the stock series is not stationary, therefore we want to make the data stationary by taking differentiation. Note that this can also be done using log transformation, but we are using differentiation here. Because if the first differentiation appears to be not stationary, we can attempt to further stablize the series by taking a second differentiation.

We plot the ACF of differentiated stock data and discover that the ACF is nicely centered around 0 and the mean appears to be constant. Therefore we conclude the first differential is stationary.

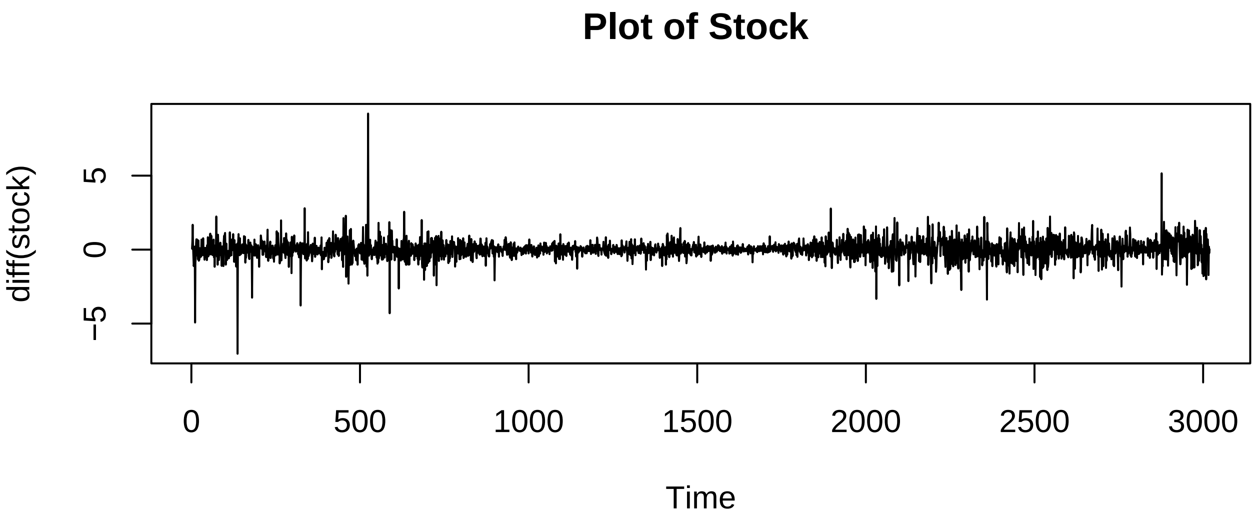


Figure 1: Plot of Stock Closing Prices

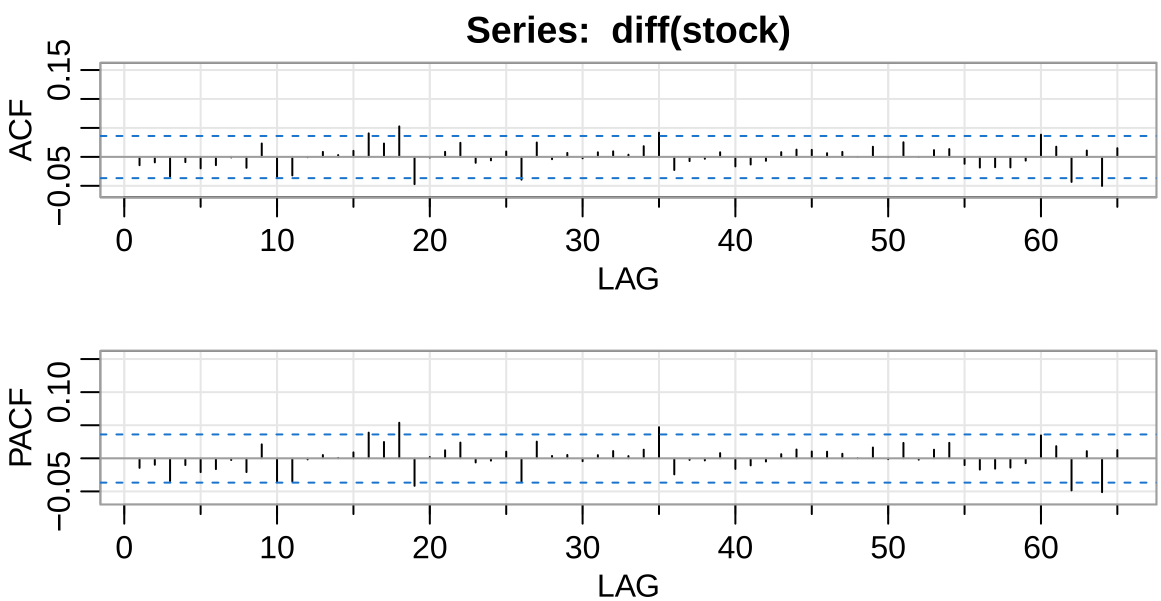


Figure 2: ACF

Autoregression Integreted Moving Average Model

First, the auto regression refers to a model where the value of the variable of interest depends on the previous terms. Moving average refers to the fact that there could be relationship between the value of the variable of interest and the residuals of previous terms. Integrated just means the data is integrated to adjust for stationary.

`auto.arima()` from `forecast` helps to choose the best ARIMA model by calculating either AIC, AICc or BIC and presenting a forecast with the most satisfying results. We are dividing the data into two groups, training and testing. We want R to generate 4 different models where we exclude different amount of observations as test data. We are choosing to exclude 600, 500, 250 and 100 observations repectively for these models. What this means is that we leave 600, 500, 250 and 100 observations behind respectively and only use the remainder to fit the model.

Results

`auto.arima()` gives 4 different models in terms of different observations dropped, with their respective forecasts:

- model1: ARIMA(2,1,2) with drift;
- model2: ARIMA(0,1,0);
- model3: ARIMA(0,1,0);
- model4: ARIMA(1,2,0).

By checking the AIC, AICc and BIC of these four models, we decide to choose the first model, ARIMA(2,1,2). Because of the seasonal component that the stock market, we fit a SARIMA(2,1,2) based on th differentiated data.This means that there are two auto regressive terms, two moving average terms and one differentiation is necessary.

Table 2: Summary of Results

headline	line1	line2	line3	line4
parameter	ar1	ar2	ma1	ma2
Estimate	-0.9334	-0.0144	-0.0831	-0.9169
p-value	0.0162	0.5489	0.8301	0.0179

Based on the significance test result, we conclude that only $\phi_1 = -0.9334$ and $\theta_2 = -0.9169$ pass the significance test with corresponding p-value of 0.0162 and 0.0179, below our threshold of $\alpha = 0.05$. Therefore the final fitted model is as following:

$$(1 + 0.9334B)\nabla x_t = (1 - 0.9169B)w_t$$

.

We see that the difference of differentiated price at a certain day is dominantly determined by the closing price of a previous day. With that model, we can predict the future stock prices of AABA.

The detailed prediction values of differentiated stock closing price, which is the rate of change of stock price at the end of a day, in shown in Table 3.

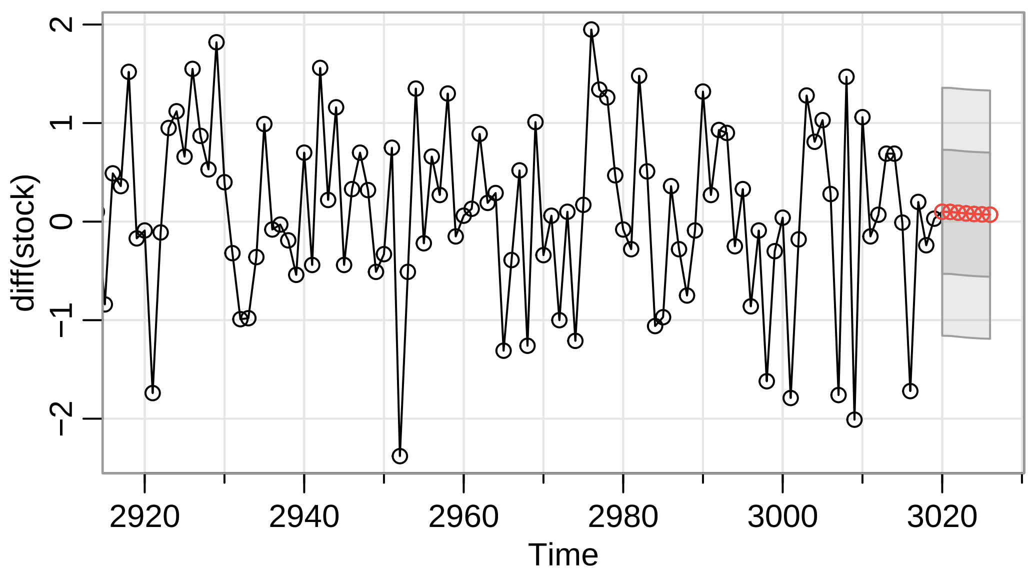


Figure 3: Week forecast

From the forecast in Fig.1, we see that based on the stock price of the previous day, the rate of change of the stock closing price continues to decrease, which means the stock prices are expected to increase but at a slower rate, given the rate of change stays above 0.

Table 3: 7 days forecast CI of AABA differentiated stock Prices-Rate of Change

headline	day1	day2	day3	day4	day5	day6	day7
Prediction	3020	3021	3022	3023	3024	3025	3026
Lower	-1.1330	-1.1347	-1.1426	-1.1509	-1.1570	-1.1612	-1.1642
Upper	1.3331	1.3318	1.3241	1.3173	1.3120	1.3082	1.3054

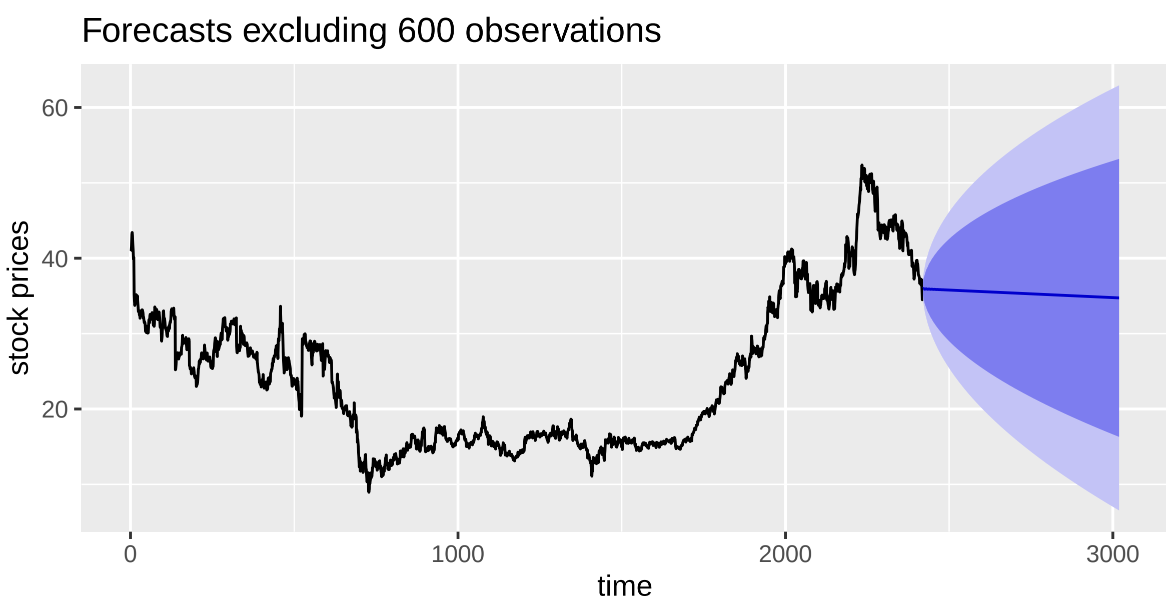


Figure 4: Forecast using first 2400 observations

Discussion

The first three predominant spectrum with the 95% confidence intervals are shown in the table below:

Table 4: First three predominantes periods of AABA Stock Price Data Set

period	spectrum	upper	lower
period1	2.7703	109.421	0.750
period2	2.5023	98.835	0.678
period3	2.3813	94.056	0.645

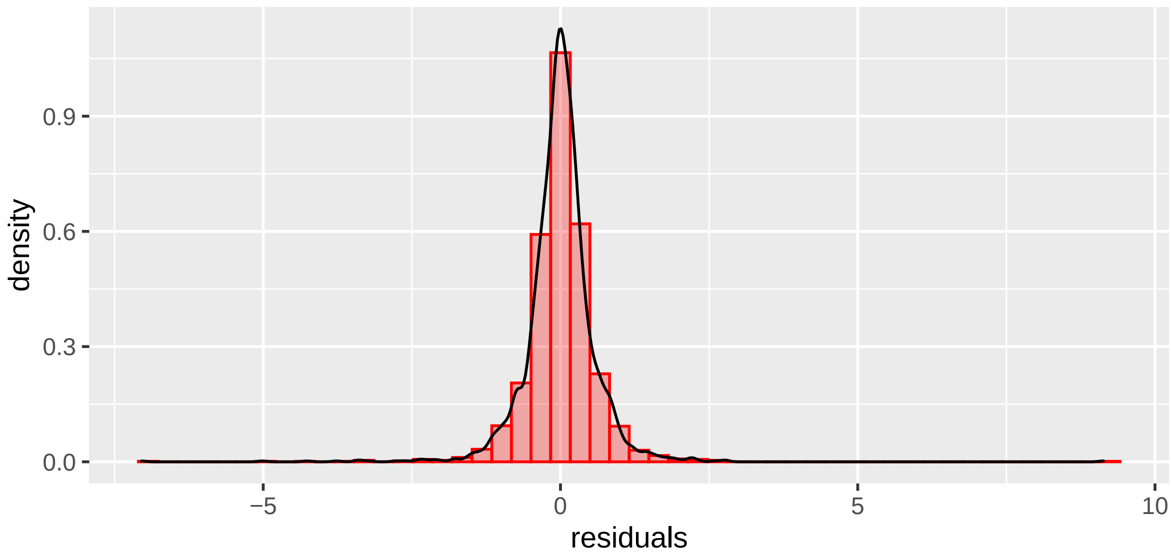


Figure 5: Residual Plot of test model

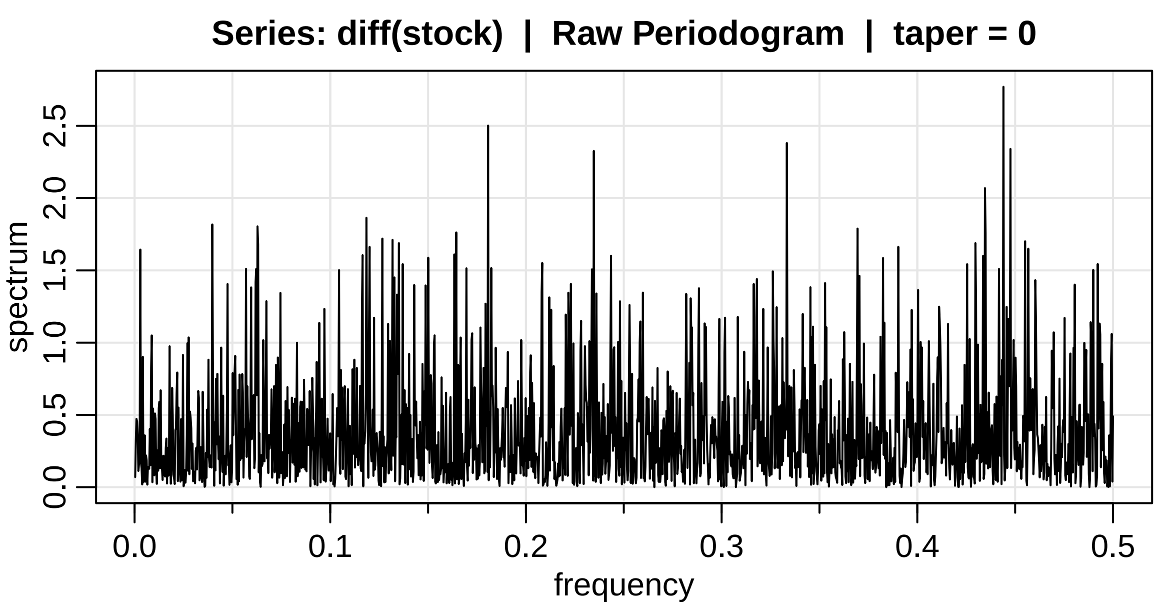


Figure 6: Spectrum

We can tell that the 95% confidence intervals for all three periods are large. For the first peak, its periodgram ordinate is 2.7703, lying the 95% confidence interval for the second peak. For the second peak, its periodgram is 2.5023 which lies in the confidence interval of the third peak. The third periodgram also lies in the confidence interval of the first two periods.

We see that the model predicts a downward trending of the first differentiation of closing Altaba stock prices based on data from 2008 to 2016. This can be explained by the model being hugely dependent on t rate of stock prices on the previous day. The downward trend of the rate of change of stocks can still be interpreted as a stock price increase because the rate of change could still be positive, but(according to the model) does not increase rapidly as it did at the previous period. In the case where the rate of change becomes negative, the stock prices will decrease. One thing with this data is that we did not deal with outliers. There seems to be quite a bit of volatilities around the first 500 days of the series also towards the end of the series.