Snap Stock Price Prediction

Nina Sun and Jeanny Zhang

June 5, 2022

1. Abstract

Snap Inc. is a tech company well known for its development of the popular multimedia instant messaging app, Snapchat. Being one of the youngest and most successful social media companies in the world right now, Snap.Inc is considered to have great potential for future growth, appealing to a lot of financial investors and tech enthusiasts. This study aims to explore the future outlook of Snap.Inc by fitting a times series model on the closing prices of the company's stock. We found that an IMA(1,1) model seems to be the most adequate model for this data, and we predicted the stock's closing prices in the next twelve days. The stock exhibits a negative trend in the next twelve days, with an overall 39.583% decrease in the closing price. Therefore, our model demonstrates that Snap.Inc will probably experience a period of financial losses after May 17, 2022.

2. Introduction

The US technology industry has always been a big player in the stock market. Recently, because of the pandemic, more and more tech companies are expanding and going public, and some of their stocks are highly expected to project considerable growth in the long run. We think it would be interesting to model a tech company's stock price and predict the company's near future outlook through forecasting its stock prices.

We want to take a closer look at Snap.Inc in this project because it is one of the most successful social media companies in the world. It develops and maintains technological products and services, with the most popular product being Snapchat which is an instant messaging application. The company was founded on September 16, 2011 and went IPO on March 2, 2017. Compared to the other tech giants such as Google and Meta, Snap.Inc is still a relatively young company with great potential for growth. Many are interested in seeing how this company will do in the future, which is what we are going to explore in this project using the stock prices of Snap.Inc.

In this report, we will use the Snap.Inc stock price data and look for a model that explains the data most. The date range of the stock price data ranges from its first day of IPO to May 17, 2022, which is the day we first obtained the data set. We will then use the best-fitting model to forecast the company's future stock prices in the next twelve days from May 18 to June 3, so that we can compare our predictions to the actual data and see how well our model performs. Limitations of this study will also be discussed at the end of this report.

3. Data

The data used in this project is obtained from the Nasdaq database (https://www.nasdaq.com/market-activity/stocks/snap/historical). The database contains all historical data of the Snap.Inc stock, including columns as date, closing price, traded volume, opening price, high price, and low price. In this study, we decided to focus on the stock's closing price. The two columns of data that we used for model-fitting are listed below.

1. Date: Everyday from 03/02/2017 to 05/17/2022 (format: MM/DD/YYYY).

2. Close.Last: Closing price of Snap.Inc stock in US dollars.

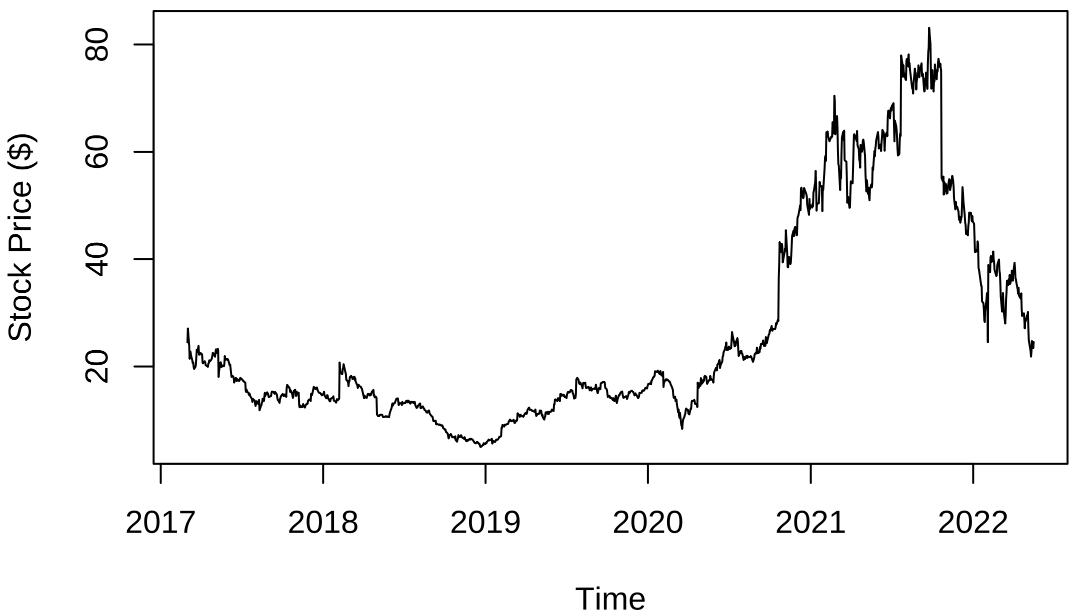
The original dataset has the date ranging from the latest to the earliest date. Therefore, we rearranged the order of the data set to make the data display in a chronological order.

4. Results

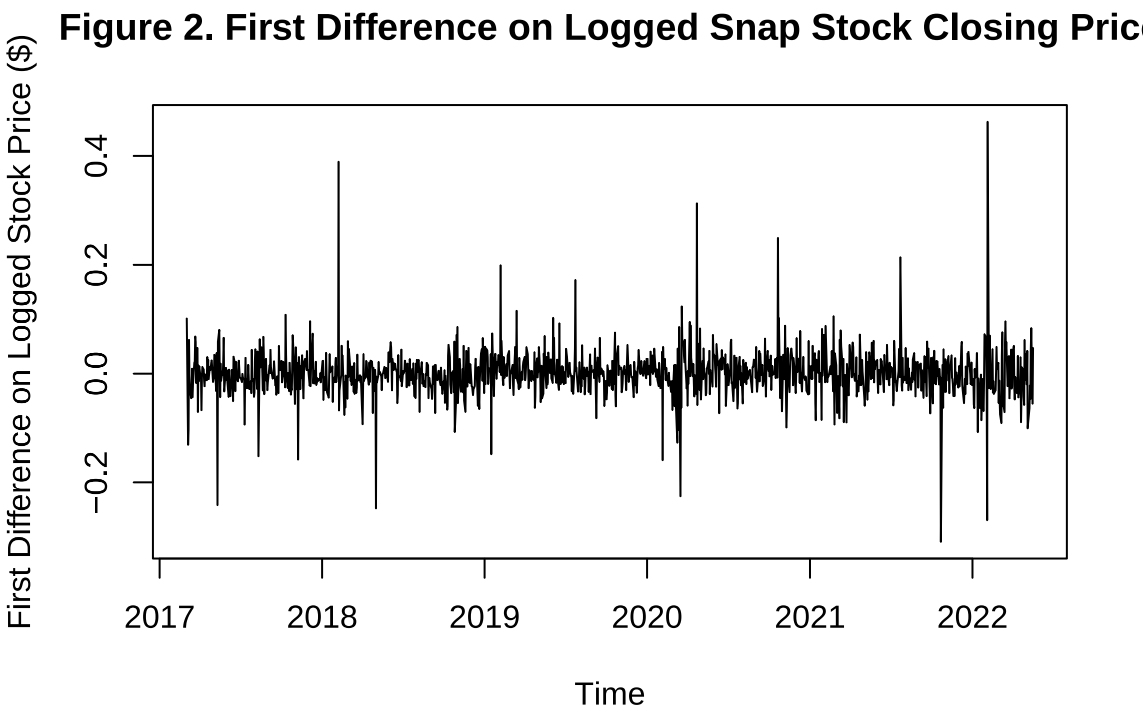
The best-fitting model that we found for the Snap.Inc stock's closing price is an IMA(1,1) model where $\theta_1 = -0.0670$ with a standard error of 0.0279.

Based on Figure 1 which shows the time series of the closing stock price, the data does not appear to be stationary due to its long meanders away from the mean and the fact that its mean and variability are not constant over time. The stock prices are relatively stable between the first day of IPO and the start of 2020, but there is a huge increase during the end of 2020 and the start of 2022.

Figure 1. Times Series of Snap Stock Closing Price



Therefore, we took the first difference of the data after a log transformation to make the time series more stationary, as shown in Figure 2.

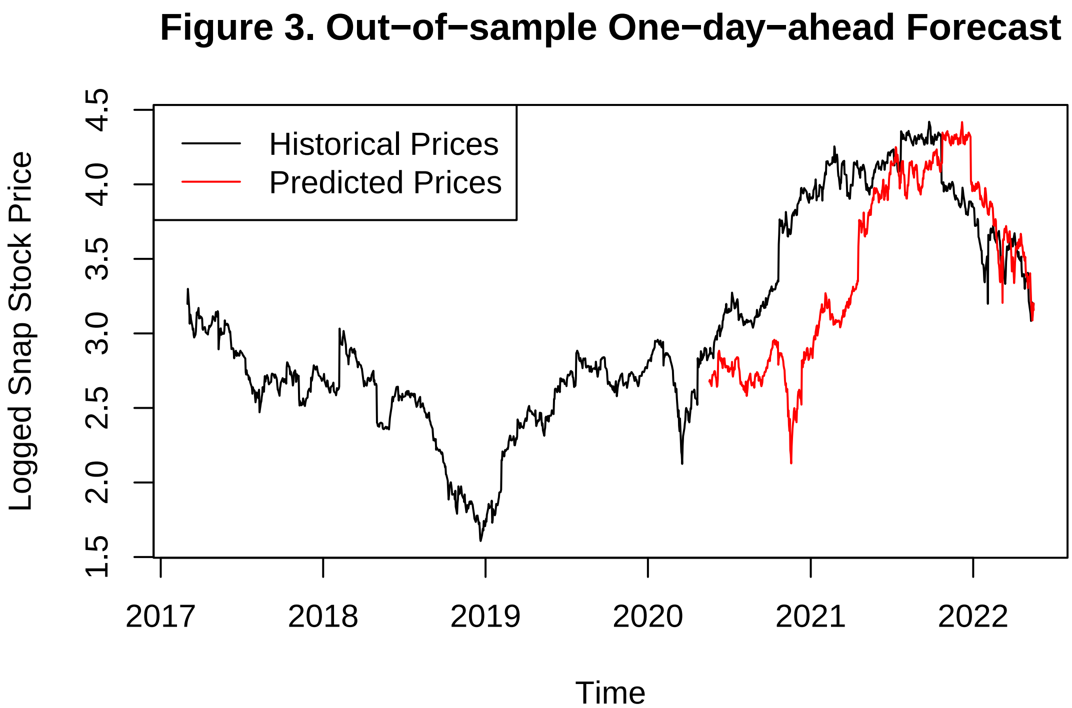


By observing the ACF and PACF plots of the logged data on the first difference, an IMA(1,1) model is suggested as both ACF and PACF demonstrate correlation at lag 1. The EACF of the data also suggests an IMA(1,1) model or ARIMA(1,1,1) model.

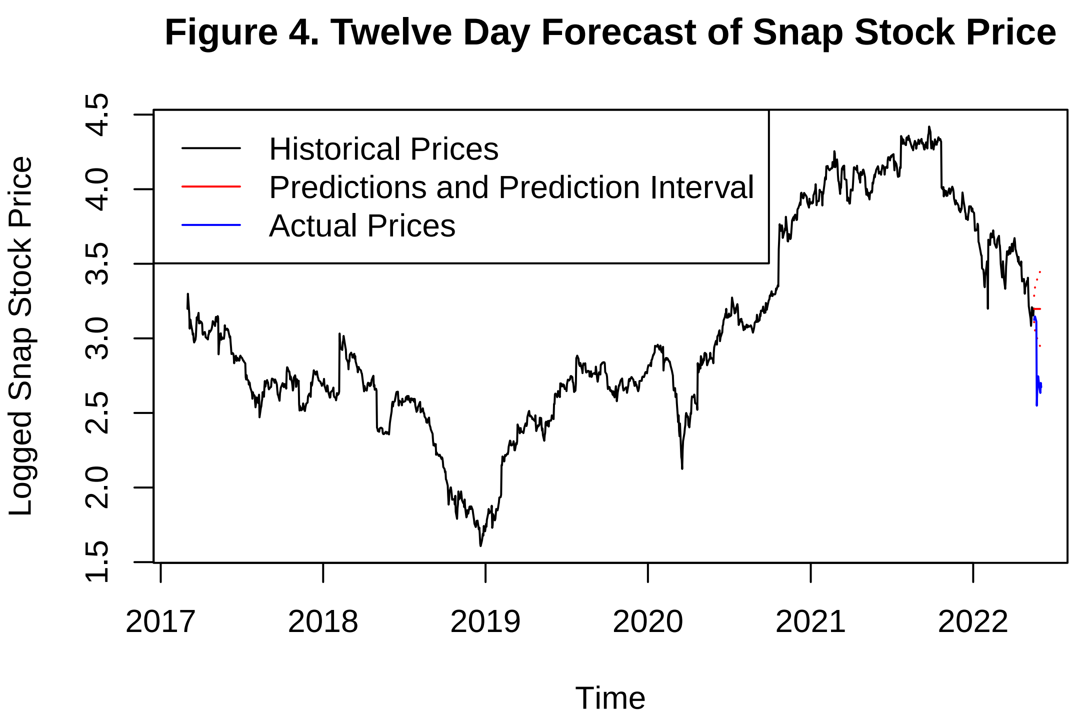
The model diagnostics of the IMA(1,1) model does not display any substantial badness in the model. The model has an AIC of -4482.416, an AICc of -4482.407, and a BIC of -4472.058. The residuals plot of the model does not show any pattern or sign of correlation. Neither of the ACF or PACF plots of the residuals indicates a significant value, meaning there's no correlation at any lag. The p-values of the Ljung-Box plot are all above the significance level, which suggests that the model is adequate.

Spectral density plots including the raw periodogram, the smoothed periodogram, and the model periodogram were plotted for the IMA(1,1) model. The model appears to fit the data reasonably well. The same previous processes were conducted to diagnose the ARIMA(1,1,1) model. The two models have almost identical periodograms but IMA(1,1) has smaller AICc, which indicates that it will produce better predictions. Thus, the IMA(1,1) model is the most adequate time series model for our data and conducting predictions.

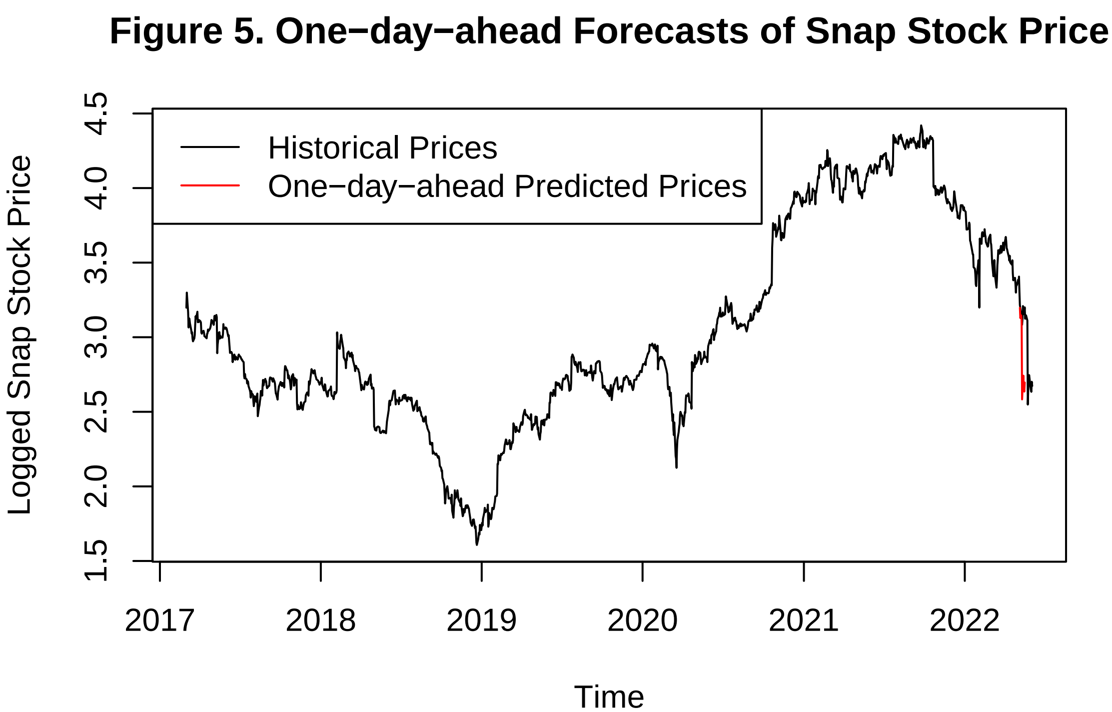
out-of-sample one-step-ahead forecast error was calculated in order to check how well the IMA(1,1) model is performing in terms of prediction. As shown in Figure 3, we used the model to conduct out-of-sample one-step-ahead predictions for the closing stock prices in the last two years. Although the predictions seem to capture most of the behaviors of the actual time series data and the resulting out of sample root mean squared error is 0.0464, there seems to be a delay in the predictions for a couple of months.



The forecasts for the closing price of Snap.Inc stock for the next twelve days are shown in Figure 4. The forecasts are constant with a value of 3.197, which is approximately $24.469 on the original scale. The forecasts are constant and the prediction intervals are increasing without bound in this case. The prediction intervals capture the actual future prices for a few days before the actual prices exceed the intervals. Overall, the twelve day forecast seems to be overestimating the closing prices of the stock.



The one-day-ahead forecast for the closing price of Snap.Inc stock for the next twelve days are shown in Figure 5. The forecasts seem to be capturing the behaviors of the actual future stock prices well and exhibit a negative trend over time. The root mean squared error is 0.171 in this case. One thing that is worth noticing is that the forecasts appear to exhibit the negative trend earlier than the reality, so there is still some misfitting of the stock prices.



5. Discussion

Based on the result section, the IMA(1,1) model is the most reasonable model according to the model specification and diagnostics. The stock exhibits a negative trend in the next twelve days. With a stock price of \$24.55 on May 17, 2022 and a predicted stock price of \$14.826 on June 3, 2022 using the one-day-ahead forecast method, the closing price of Snap.Inc stock will see an overall 39.583% decrease. This result is very similar to the trend of actual stock prices. The actual closing stock price on June 3, 2022 is \$14.49, which suggests an overall 40.954% decrease. Therefore, our model correctly predicts that Snap.Inc will experience a period of stock price crashes and financial losses after May 17, 2022.

There also exists some limitations that affect our model fitting and prediction. First, stock prices vary each second, and many professional organizations are trying to establish models to predict future prices in order to gain profit. Even though the IMA(1,1) model looks adequate enough in our diagnoses, our current knowledge doesn't support us well enough to apply a more complex model to better predict the price variation. Therefore, there still might be model misfits that influence our forecasting. Second, outside factors such as announcements regarding the company revenue, employment adjustments, and company reputation could also heavily influence the stock price. However, time series models that we have learned so far fit better on data that vary naturally. Stock prices can be manipulated by big financial organizations, which means that they do not vary naturally in some circumstances. Therefore, trying to fit an integrated first-moving average model on unnatural data might not be the best way to explain the changes in stock price.

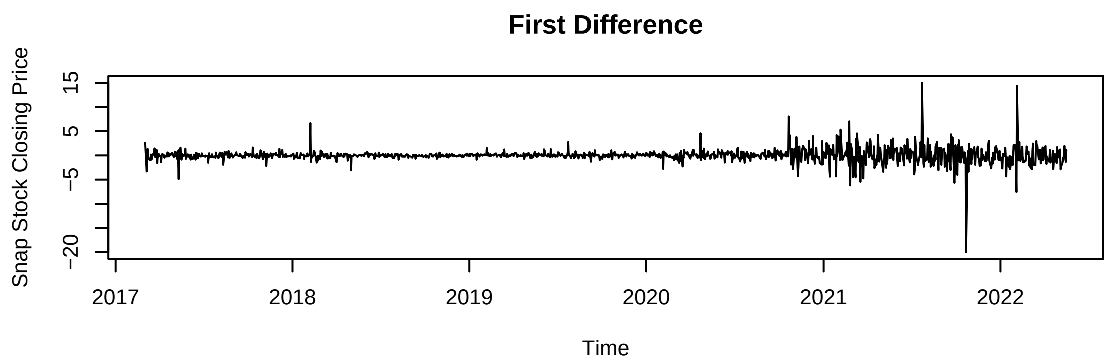
Further research can be involved to investigate the effect brought by the pandemic, as Snap Inc.'s price rose drastically at the start of the pandemic and later dropped back. We can explore how the pandemic changes the best-fitting time series model. Meanwhile, we can also look closer to the data set by looking at the price variation within seconds instead of days, and compare the best-fitting model at different time units.

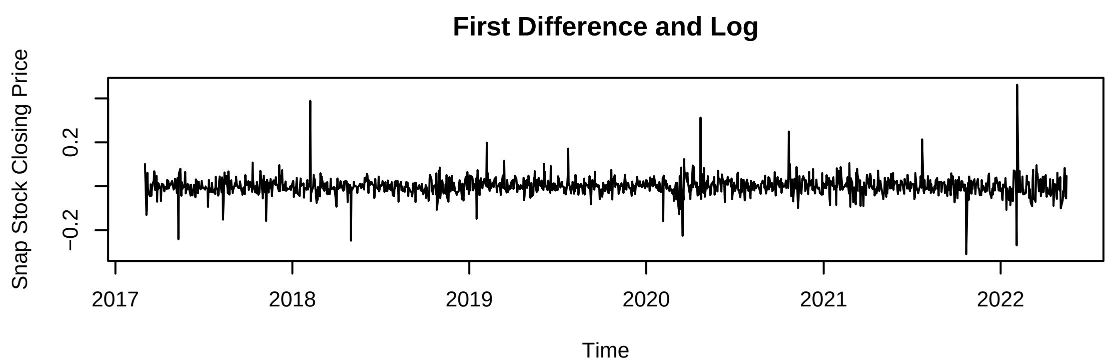
6. References

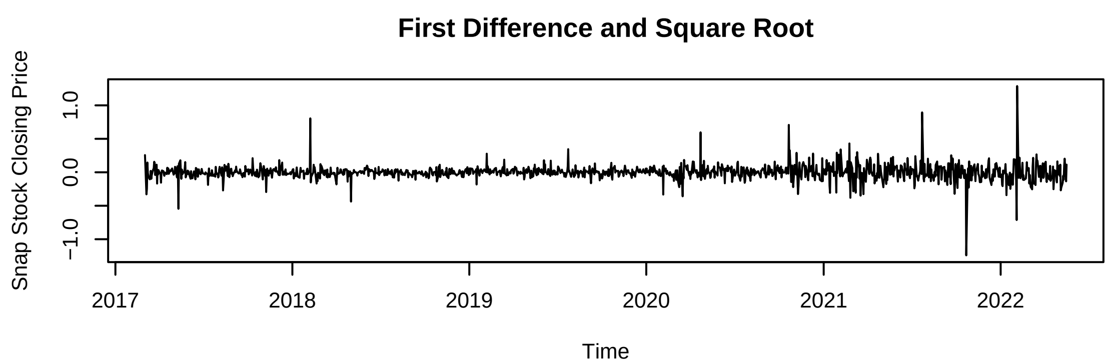
"Snap Inc. Class A Common Stock (SNAP) Historical Data | Nasdaq." Accessed June 6, 2022. https://www.nasdaq.com/market-activity/stocks/snap/historical.

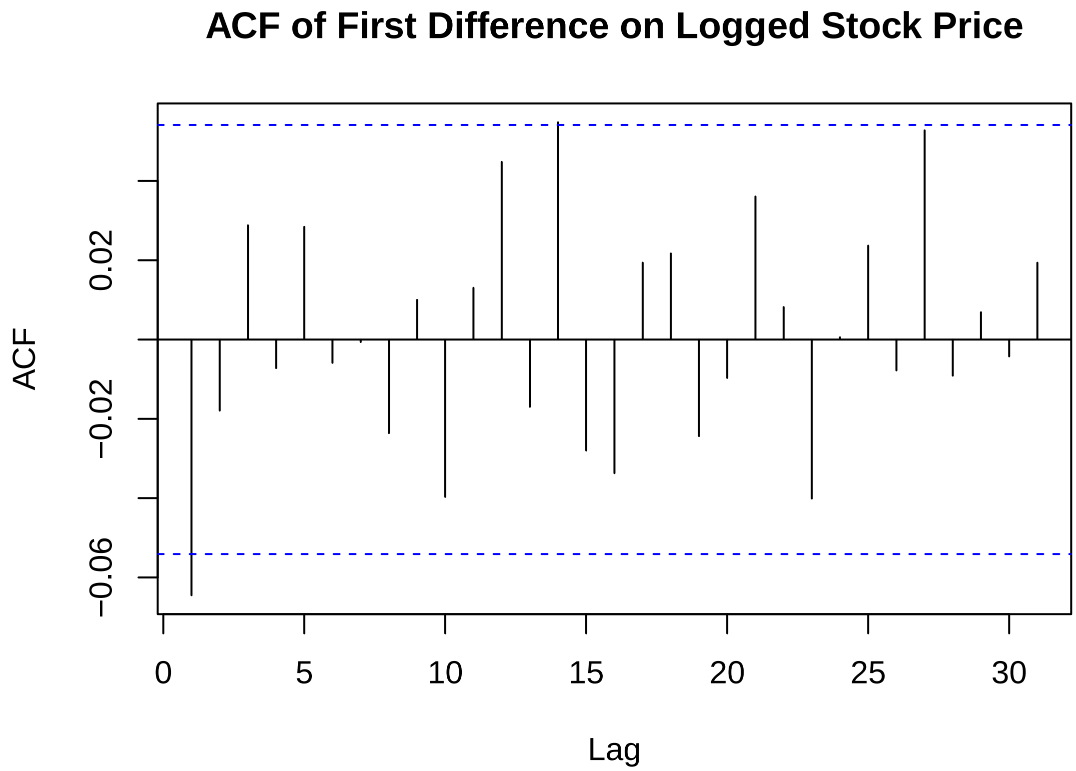
"Snap Inc." In Wikipedia, May 11, 2022. https://en.wikipedia.org/w/index.php?title=Snap_Inc.&oldid=1087326997.

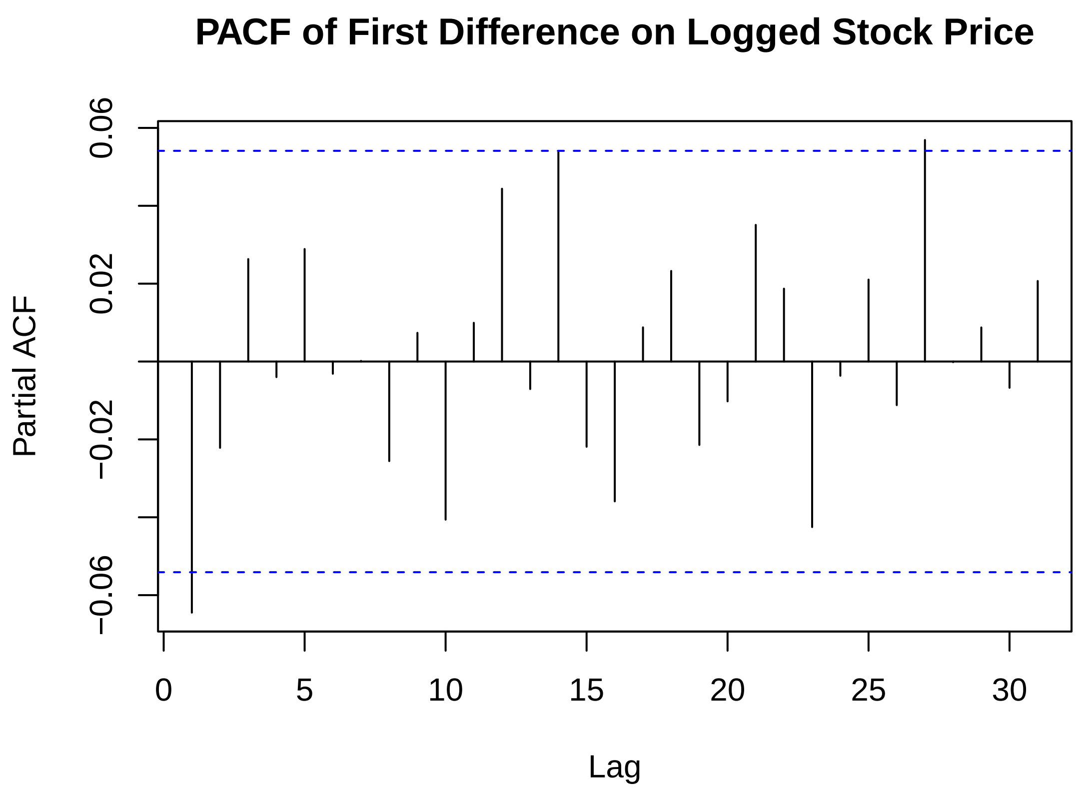
7. Appendix











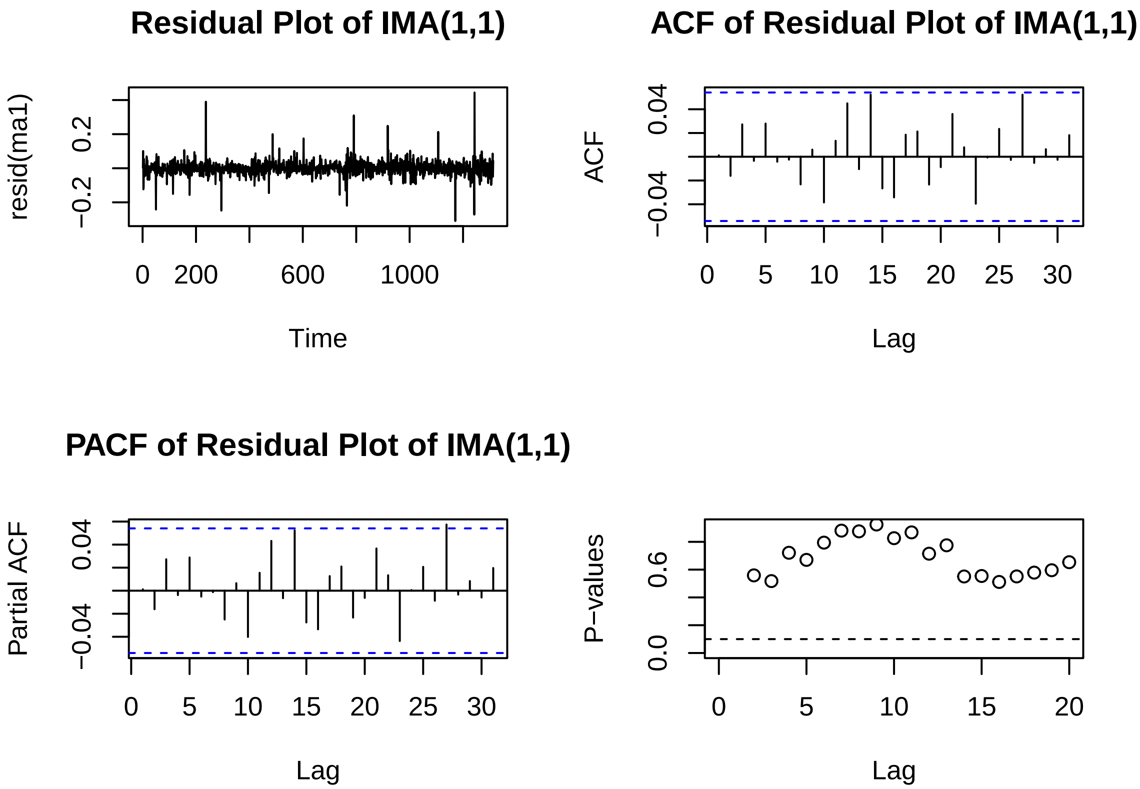

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x o o o o o o o o o o  o  o  o
## 1 x o o o o o o o o o o  o  o  o
## 2 x x o o o o o o o o o  o  o  o
```

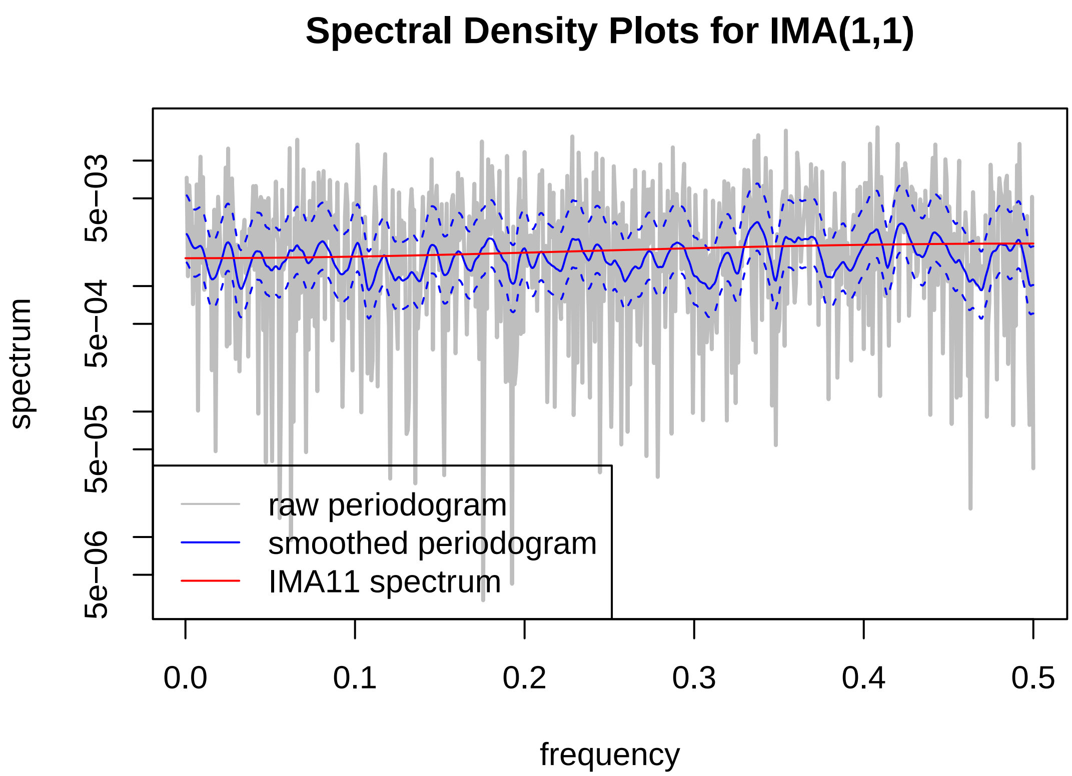
```
## 3 x x o o o o o o o o o  o  o  o
## 4 x x x x o o o o o o o  o  o  o
## 5 x o x x x o o o o o o  o  o  o
## 6 o o x x x o o o o o o  o  o  o
## 7 o o x x x o o o o o o  o  o  o
```

```
## [1] -4482.416
```

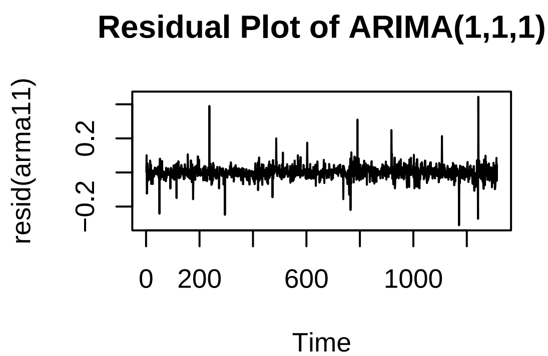
```
## [1] -4482.407
```

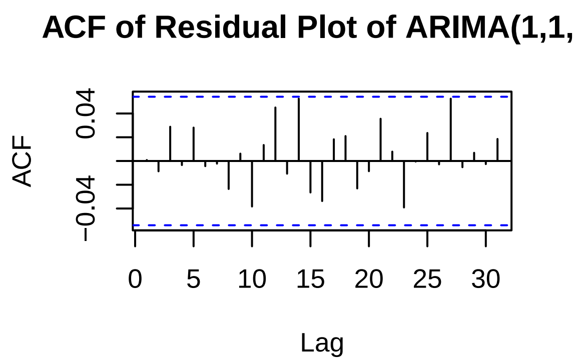
```
## [1] -4472.058
```

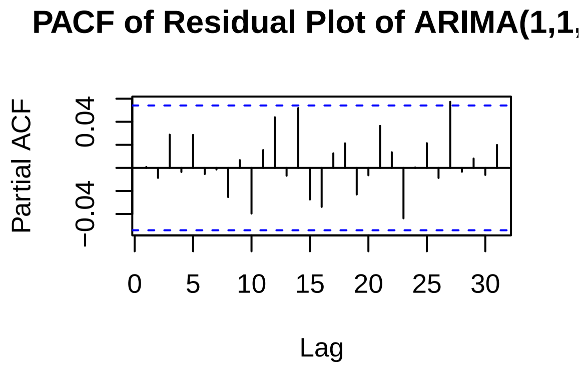


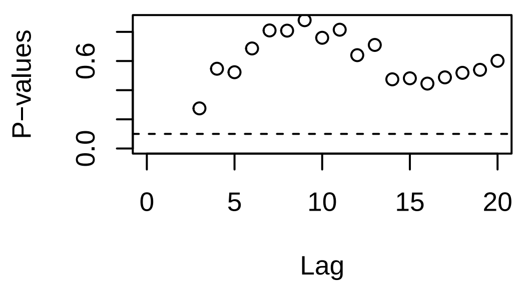


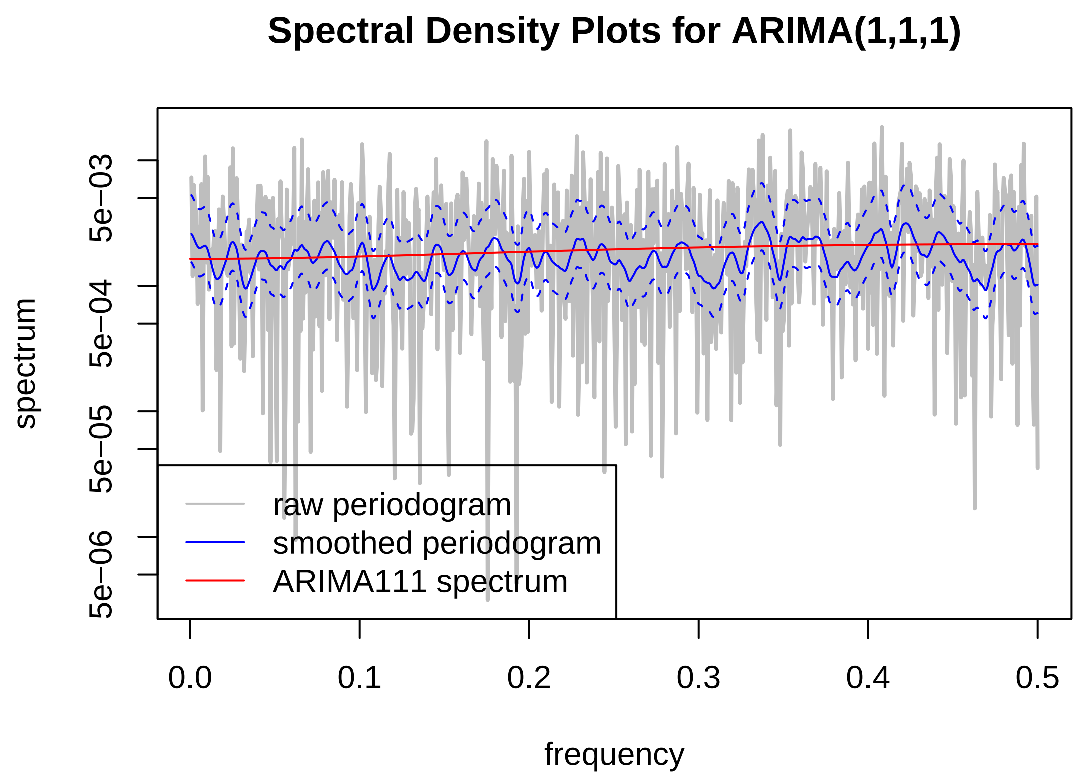

```
## [1] -4480.551
## [1] -4480.532
## [1] -4465.013
```











8. Code Appendix

Code in Report

```r
# load libraries
library(tidyverse)
library(TSA)
library(lubridate)
library(kableExtra)

# load data set
df <- read.csv("SnapStock_HistoricalData.csv")
df$Date2 <- mdy(df$Date) # format date
df$`Close.Last2` <- as.numeric(gsub("\\$", "", df$Close.Last)) # remove dollar sign from price
df <- df %>% arrange(desc(row_number())) # reverse the rows' order

# plot time series of Snap stock closing price
plot(y = df$Close.Last2, x = df$Date2, xlab = "Time", ylab = "Stock Price ($)", main = "Figure 1. Times
# plot first difference on logged snap stock closing price
plot(y = diff(log(df$Close.Last2)), x = df$Date2[2:1313],  xlab = "Time", ylab = "First Difference on L

# calculate root mean squared errors of IMA(1,1) in the last 2 years
ma1 <- arima(log(df$Close.Last2), order = c(0,1,1))
n <- dim(df)[1] # total sample size
start_time <- dim(df)[1]-365*2 # start process two years from now
numPreds <- 365*2
preds_ma1 <- numeric(numPreds)
errors_ma1 <- numeric(numPreds)
```

```
for(t in 1:numPreds){
  # fit MA(1) to first block of data
  ma1_t <- arima(log(df$Close.Last2[1:(start_time + t - 1)]), c(0,1,1))

  # predict next day's value from model
  preds_ma1[t] <- predict(ma1_t)$pred

  # calculate prediction error
  errors_ma1[t] <- log(df$Close.Last2[start_time + t]) - preds_ma1[t]
}

# calculate root mean squared errors
sqrt(mean(errors_ma1^2))

# plot one-day-ahead out-of-sample forecast of IMA(1,1) for the last 2 years
plot(y = log(df$Close.Last2), x = df$Date2, type = 'l', xlab = "Time", ylab = "Logged Snap Stock Closing
# add sequential predictions
lines(y = preds_ma1, x = seq.Date(from = max(df$Date2)-365*2+1, to = max(df$Date2), by = 'days'), col =
legend("topleft", col = c("black", "red"),
       legend = c("Historical Prices", "Predicted Prices"), lty = 1)

# load data set with stock prices for the next 10 days
df2 <- read.csv("SnapStock_HistoricalData2.csv")
df2$Date2 <- mdy(df2$Date)
df2$`Close.Last2` <- as.numeric(gsub("\\$", "", df2$Close.Last))
df2 <- df2 %>% arrange(desc(row_number()))
# join the two datasets
df2 <- df %>%
  full_join(df2, by = c("Date2", "Close.Last2")) %>%
  select(Date2, Close.Last2)

# predict stock prices for the next 12 days
preds <- predict(ma1, n.ahead = 12)

# plot predictions and prediction interval
plot(y = log(df2$Close.Last2)[1:1313], x = df$Date2, type = 'l', xlab = "Time", ylab = "Logged Snap Sto
lines(y = preds$pred, x = df2$Date2[1314:1325], lty = 'dashed', col = 'red')
lines(y = preds$pred + 2*preds$se, x = df2$Date2[1314:1325], lty = 'dotted', col = 'red')
lines(y = preds$pred - 2*preds$se, x = df2$Date2[1314:1325], lty = 'dotted', col = 'red')
lines(y = log(df2$Close.Last2)[1314:1325], x = df2$Date2[1314:1325], col = 'blue')
legend("topleft", col = c("black", "red", "blue"),
       legend = c("Historical Prices", "Predictions and Prediction Interval", "Actual Prices"), lty = 1

# calculate root mean squared errors of IMA(1,1) for the next 12 days
n <- dim(df2)[1] # total sample size
start_time <- dim(df2)[1]-12 # start process 12 days from now
numPreds <- 12
preds_ma1 <- numeric(numPreds)
errors_ma1 <- numeric(numPreds)

for(t in 1:numPreds){
  # fit MA(1) to first block of data
  ma1_t <- arima(log(df2$Close.Last2[1:(start_time + t - 1)]), c(0,1,1))
```

```
  # predict next day's value from model
  preds_ma1[t] <- predict(ma1_t)$pred

  # calculate prediction error
  errors_ma1[t] <- log(df2$Close.Last2[start_time + t]) - preds_ma1[t]
}

# calculate root mean squared errors
sqrt(mean(errors_ma1^2))

# plot one-day-ahead out-of-sample forecast of IMA(1,1) for the next 12 days
plot(y = log(df2$Close.Last2), x = df2$Date2, type = 'l', xlab = "Time", ylab = "Logged Snap Stock Clos
# add sequential predictions
lines(y = preds_ma1, x = seq.Date(from = max(df$Date2)-11, to = max(df$Date2), by = 'days'), col = 'red
legend("topleft", col = c("black", "red"),
       legend = c("Historical Prices", "One-day-ahead Predicted Prices"), lty = 1)
```

Code in Appendix

```
# try to make the data more stationary
par(mfrow = c(3,1))
plot(y = diff(df$Close.Last2), x = df$Date2[2:1313],  xlab = "Time", ylab = "Snap Stock Closing Price",
plot(y = diff(log(df$Close.Last2)), x = df$Date2[2:1313],  xlab = "Time", ylab = "Snap Stock Closing Pr
plot(y = diff(sqrt(df$Close.Last2)), x = df$Date2[2:1313],  xlab = "Time", ylab = "Snap Stock Closing P

# model specification
acf(diff(log(df$Close.Last2)), main = "ACF of First Difference on Logged Stock Price")
pacf(diff(log(df$Close.Last2)), main = "PACF of First Difference on Logged Stock Price")
eacf(diff(log(df$Close.Last2)))

AICc <- function(model) {
  k <- length(coef(model))
  n <- nobs(model)
  return(AIC(model) + 2*(k+1)*(k+2) / (n-k-2))
}

# IMA(1,1) diagnostics
ma1 <- arima(log(df$Close.Last2), order = c(0,1,1))
AIC(ma1)
AICc(ma1)
BIC(ma1)
par(mfrow = c(2,2))
plot(resid(ma1), type = "l", main="Residual Plot of IMA(1,1)")
acf(resid(ma1), main = "ACF of Residual Plot of IMA(1,1)")
pacf(resid(ma1), main = "PACF of Residual Plot of IMA(1,1)")
source("http://people.carleton.edu/~apoppick/ClassData/LjungBoxPlot.R")
LjungBoxPlot(ma1)

# spectral density plots for IMA(1,1)
S_hat <- spec.pgram(diff(log(df$Close.Last2)), detrend = FALSE, demean = TRUE, plot = FALSE) #raw perio
S_bar <- spec.pgram(diff(log(df$Close.Last2)), kernel("modified.daniell", c(4, 4)), detrend = FALSE, de
S_bar_CI <- cbind(S_bar$spec*S_bar$df / qchisq(0.975, df = S_bar$df), S_bar$spec*S_bar$df / qchisq(0.025
```

```
ma1_spec <- ARMAspec(model = list(ma = ma1$model$theta,sigma2 = ma1$sigma2), plot = FALSE)
plot(S_hat$spec*exp(-digamma(1)) ~ S_hat$freq, type='l', col = 'gray', xlab = "frequency", ylab = "spec
lines(S_bar$spec ~ S_bar$freq, col = 'blue')
lines(S_bar_CI[,1] ~ S_bar$freq, col = 'blue', lty = 'dashed')
lines(S_bar_CI[,2] ~ S_bar$freq, col = 'blue', lty = 'dashed')
lines(ma1_spec$spec ~ ma1_spec$freq, col = 'red')
legend("bottomleft", col = c("gray", "blue", "red"), lty = 1,
legend = c("raw periodogram", "smoothed periodogram", "IMA11 spectrum"))

# ARIMA(1,1,1) diagnostics
arma11 <- arima(log(df$Close.Last2), order = c(1,1,1))
AIC(arma11)
AICc(arma11)
BIC(arma11)
par(mfrow = c(2,2))
plot(resid(arma11), type = "l", main = "Residual Plot of ARIMA(1,1,1)")
acf(resid(arma11), main = "ACF of Residual Plot of ARIMA(1,1,1)")
pacf(resid(arma11), main = "PACF of Residual Plot of ARIMA(1,1,1)")
LjungBoxPlot(arma11)

# spectral density plots for ARIMA(1,1,1)
S_hat <- spec.pgram(diff(log(df$Close.Last2)), detrend = FALSE, demean = TRUE, plot = FALSE) #raw perio
S_bar <- spec.pgram(diff(log(df$Close.Last2)), kernel("modified.daniell", c(4, 4)), detrend = FALSE, de
S_bar_CI <- cbind(S_bar$spec*S_bar$df / qchisq(0.975, df = S_bar$df), S_bar$spec*S_bar$df / qchisq(0.02
arma11_spec <- ARMAspec(model = list(ar = arma11$model$phi, ma =arma11$model$theta, sigma2 = arma11$sig
plot(S_hat$spec*exp(-digamma(1)) ~ S_hat$freq, type='l', col = 'gray', xlab = "frequency", ylab = "spec
lines(S_bar$spec ~ S_bar$freq, col = 'blue')
lines(S_bar_CI[,1] ~ S_bar$freq, col = 'blue', lty = 'dashed')
lines(S_bar_CI[,2] ~ S_bar$freq, col = 'blue', lty = 'dashed')
lines(arma11_spec$spec ~ arma11_spec$freq, col = 'red')
legend("bottomleft", col = c("gray", "blue", "red"), lty = 1,
legend = c("raw periodogram", "smoothed periodogram", "ARIMA111 spectrum"))
```