**Task 1 Efficient Market Hypothesis**

**Step 1: Obtain the Data**

**INTC over a span of 15 years (08/31/2005 to 08/31/2019)**

You can compute the simple return as:

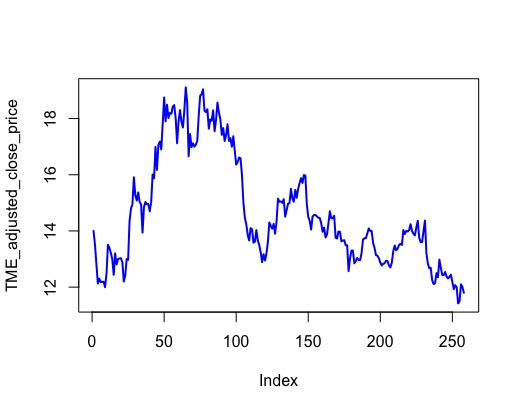
 Rt= [(Yt-Yt-1)/ Yt-1]\*100

 where is return at time t, Yt is the Adjusted close price at time t and Yt-1 is the Adjusted close price at time t-1

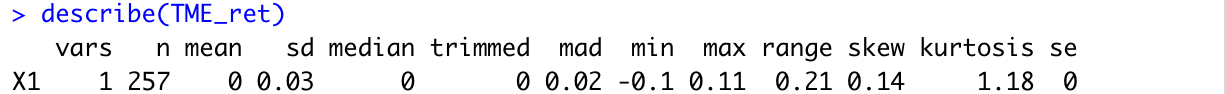
**Step 2: Interpretation on the Test of Weak Form Efficiency**

**1) Descriptive Statistics and return distributions**

Begin by analyzing the data using the summary statistics (e.g., mean, median, max, min, skewness, kurtosis etc.) in a Table.  Provide a few (may be 1 or 2) plots of time series of returns of the data and a histogram of the return distribution for your stocks and indices. Using appropriate wording you should comment on the statistical and economic (if any) interpretation of your data.

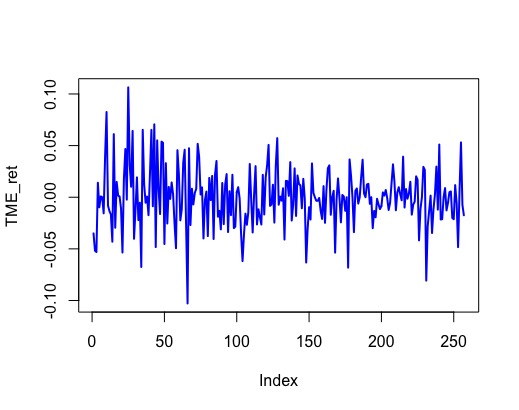
According to the plot below, for TME’s **stock price**, it reaches peak at $19, but since then, it starts to have downward trend despite small upward retracements at time. Variance of stock price also increases as in the right half of the figure, there is more ups and downs.

Comment on the statistical and economic interpretation on return:

According to summary of statistic of **TME’s return,**

1.its max is 11%, its min is -10%, its median is 0, mean is 0, and its variance is 1.

2.its skewness is 0.14, which means the data is skewed to the right, and its kurtosis is 1.18, which means its distribution is leptokurtic(kurtosis>0)—producing more outliers than the normal distribution. The return data is plotted in below.



According to the plot of **TME’s return** above,

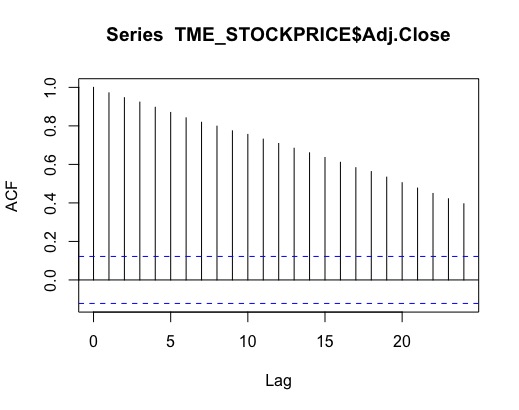
3.the variance of the data is larger in the first 100 days but becomes smaller after then

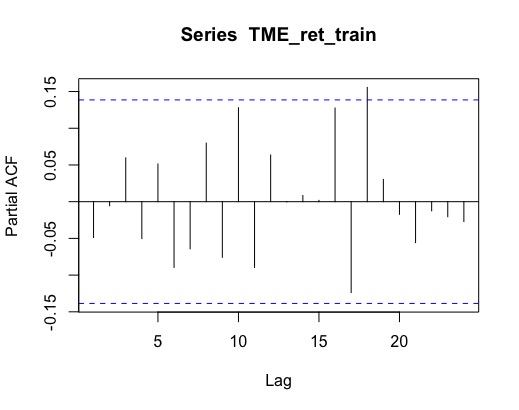
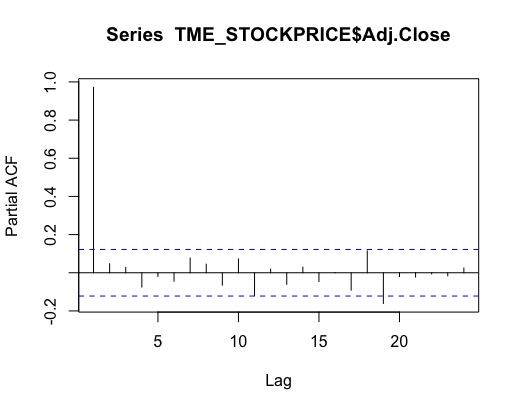
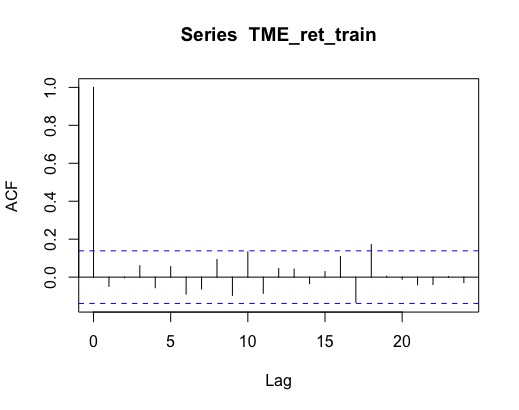
4.there seems to be a seasonal component, which has a cycle less than 12 months

5.the data looks non-stationary, so we need to take the difference of the dataset

**Autoregressive(AR) Model ACF and PACF plots:**

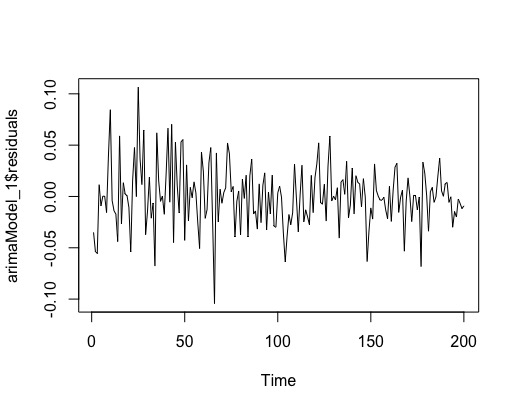
ACF (Auto-Correlation Function) is the correlation between the observation at the current time spot and the observations at pervious time spot. It is used to measure Moving Average model.

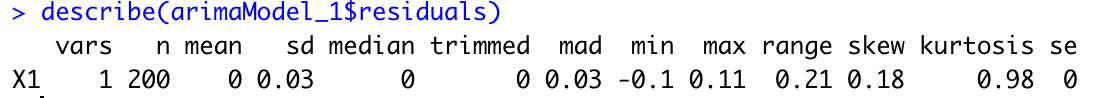
PACF (Partial Auto-Correlation Function) is the correlation between observations at two-time spots given that we consider both observations are correlated to observation at other time spots. It is used for Autoregressive model, considering the real correlation between two days.

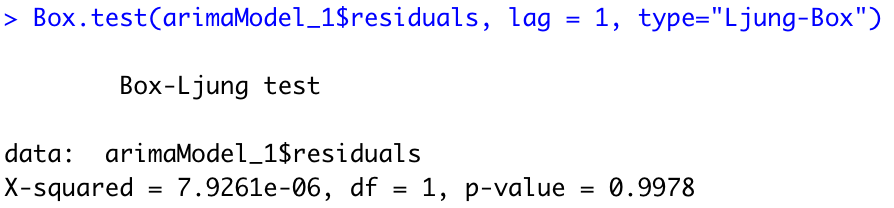
To determine which one, we need to use for time series analysis, AR model or MA model to look at if there is an obvious trend in the stock price dataset. I graphed the ACF and PACF. The blue horizontal line is the significant threshold.

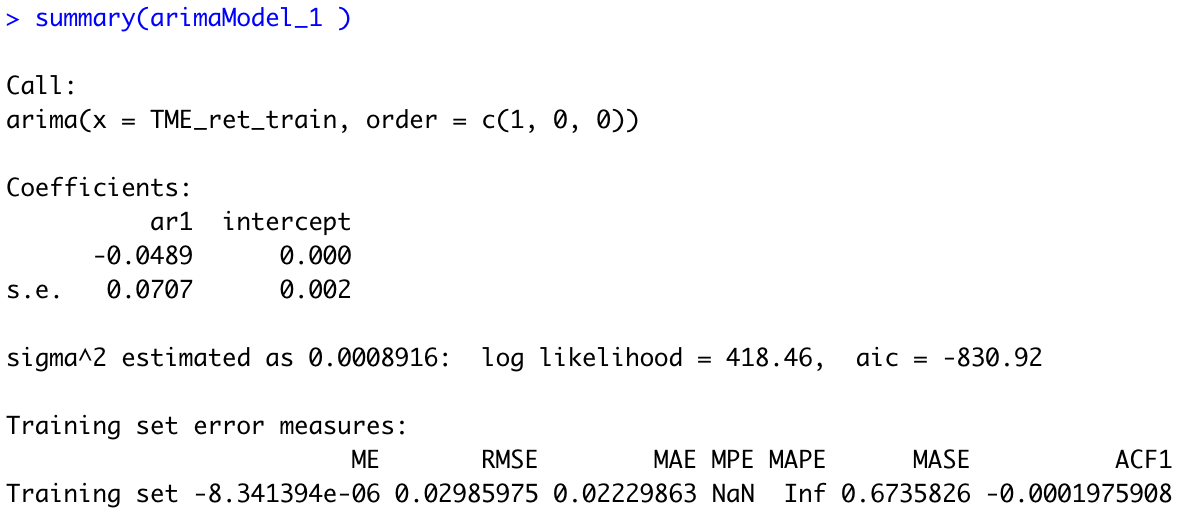
Auto Correlation Function(ACF) of TME’s adjusted close price decrease with the increase of lag, and it’s almost linear. As ACF measures the correlation between the observation at the current time spot and the observations at pervious time spot, it shows that **for all different levels of lags, the correlation between consecutive time spots are above the significant threshold**, as indicated by the blue lines. Whereas, PACF is only significant for three days. Thus, we should use AR(1 model)

#Testing the return data in your sample follow a random walk using the AR (1) model: Rt= σ+ β1 Rt-1+ ℇtwhere the dependent variable Rt is the return for the time t, and the independent variable Rt-1is the return lagged one period for time t-1#

This is result of residual by **AR model1(p=1, q=0, d=0)**:

According to test statistics, the residual from Arima model has zero mean

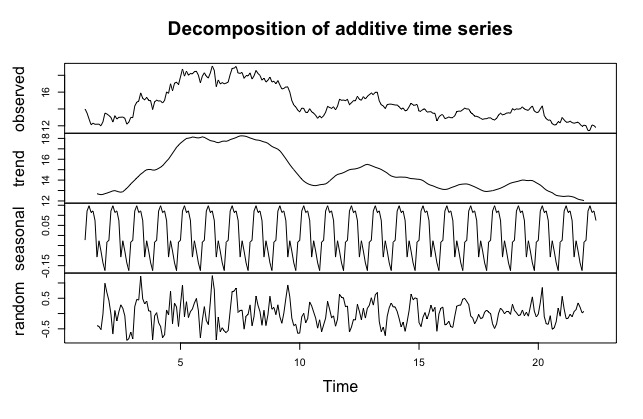
According to the Ljung-Box Test as shown below, we cannot refute the null hypothesis that the two time series isn’t autocorrelated. Thus, it is highly probable that the auto-correlation between 1-day lag return and return data is almost 0.

The summary of the model is given by the following:

Because forecasting based on model that does NOT take account into seasonality results in a completely-off prediction, we will consider the updated SARIMA model.

**3)Day of the Week Effect**

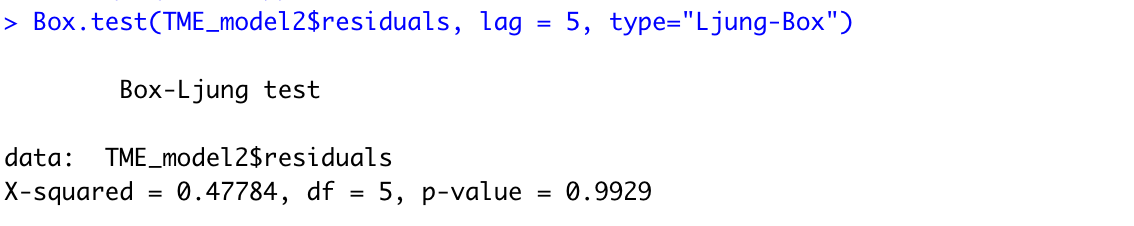
Data shows seasonality effects over day of the week

According to the graph below that decomposes additive time series, we can see the data has seasonal component:

We use the SARIMA model: Arima(1,0,18)(1,0,18)

**Interpret your response to the follow questions:**

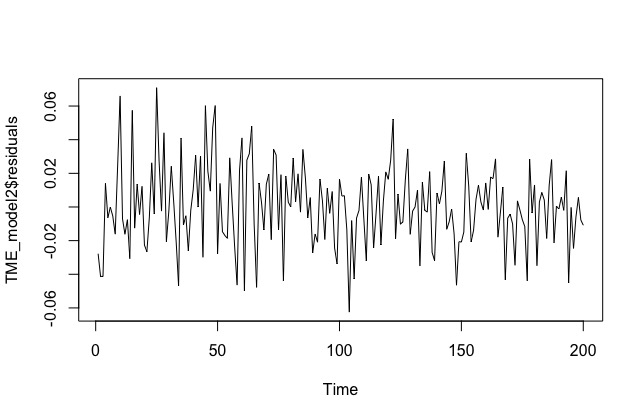
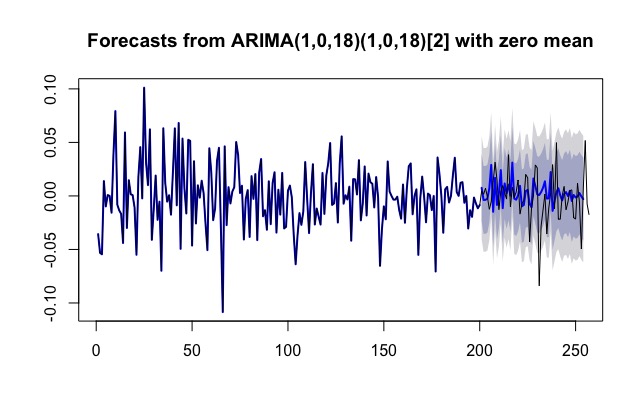
(i)            **Goodness of fit:**how well does the model describe the relationship between the variables?{ Look at the adjusted R2value}.

The Box-Ljung test tests the lack of fit of a time series model. Based on this test, we conclude that there is enough evidence to claim that the residuals are random (p-value=0.9929)

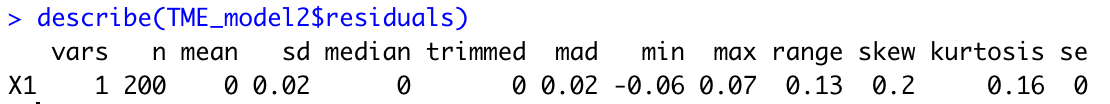
(ii)           **Overall model significance:**is there a linear relationship between all X variables taken together with Y ? is the ration of explained and unexplained variances greater than 1?

Based on Box-Ljung test, because p value > 0.05 there are no significant autocorrelations between successive forecasting errors

(iii)          Do the assumptions underlying the model (e.g., OLS regression) hold?

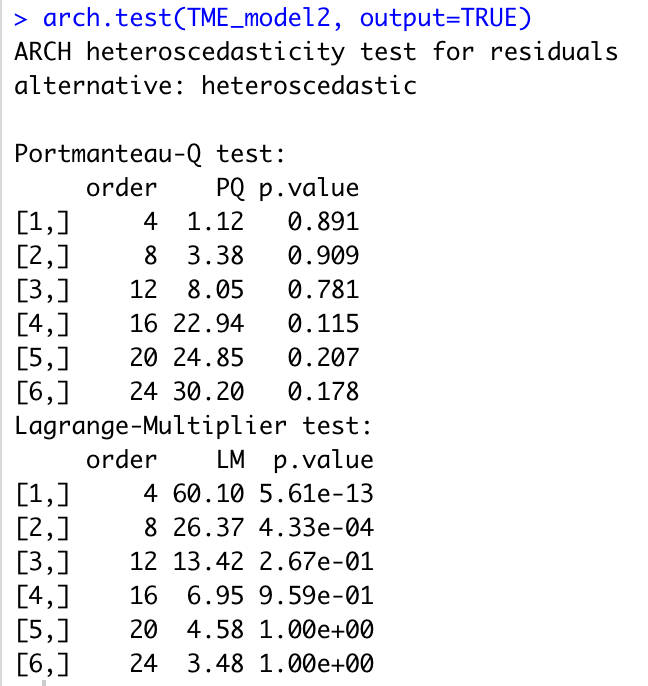
(a)  SAVE predicted and residual values

The above left is predicted and above right is residual

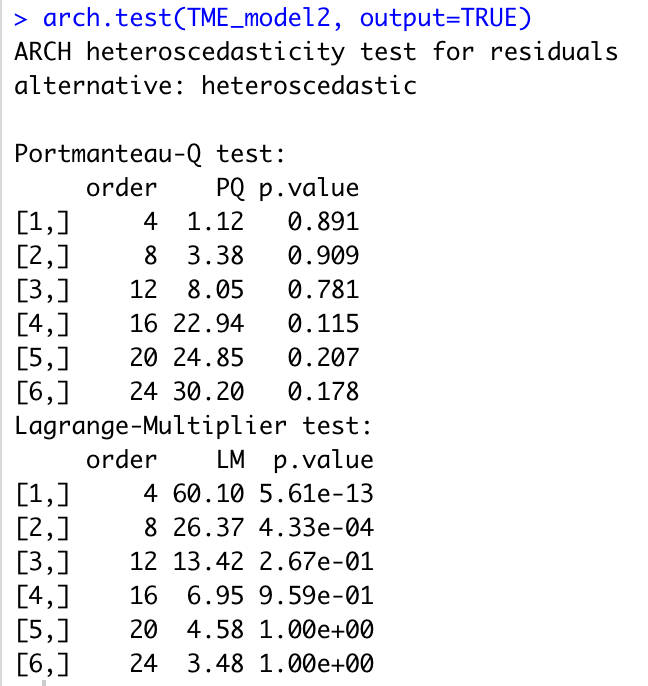
(b)  E(u)=0: the errors have **zero mean**{Look at the `residual statistics` table)

Thus, zero mean assumption holds

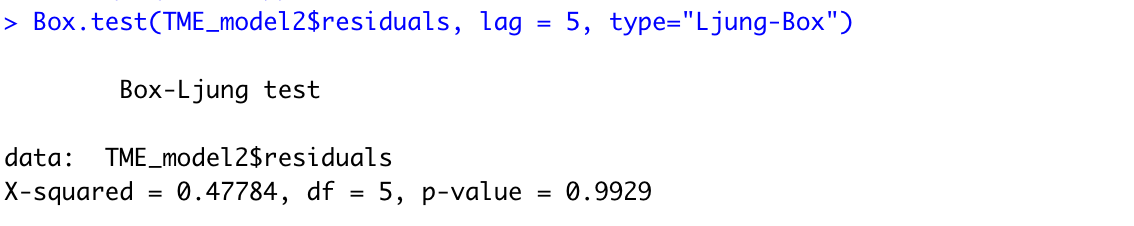
(c)   Cov(ui,uj)=0 : The errors are linearly independent of one another e.g., **no autocorrelation {**Look at the Durbin-Watson/DW value from `Model Summary` table. If the DW statistic is close to 2, we CAN`T reject ` H0 : No serial correlation`. This means the errors are independent }

From the summary statistic, there is no autocorrelation based on the Lagrange multiplier test.(we cannot not use DW test because Arima model2 is NOT linear)

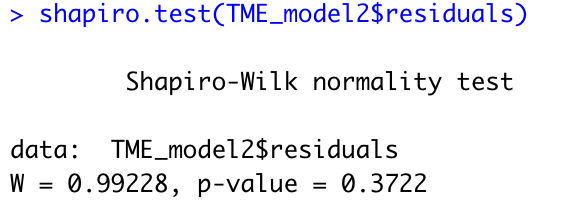
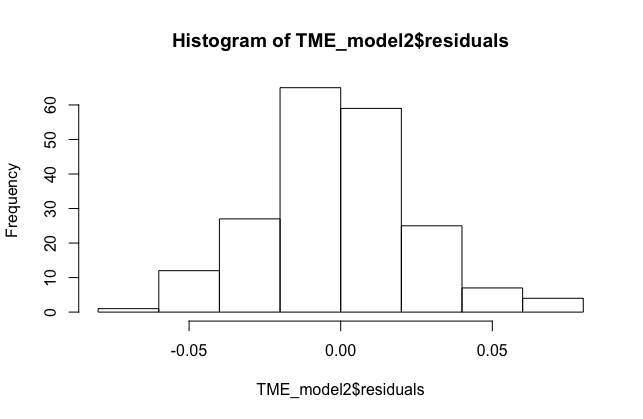
(d)  Var(u)=σ2: The Variance of the errors is constant/Homoskedastic/ equal e.g, **No Heteroskedastity**{ Draw scatter plot: Y= Standardized residuals and X= Dependent Variable. IF you see that the errors increase or decrease with an increase of the dependent variable, you`ve got heteroskedascity problem}.

Because p-value>0.05=>cannot refuse null hypothesis that there is no heteroscedasticity, the variance of the errors is constant(homoscedestic) according to the arch test.

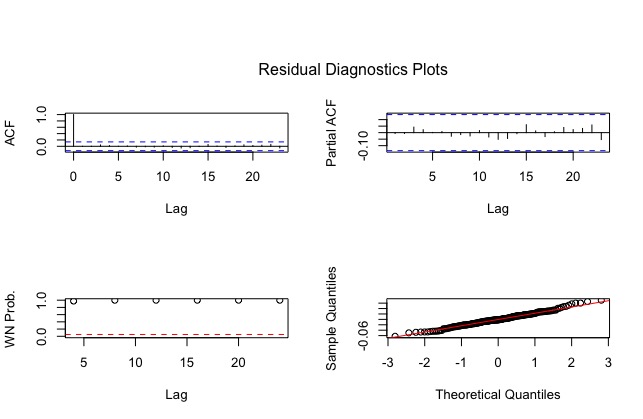
(e)  Cov(u,x) =0; There is no relationship between the error and corresponding X variable {Run correlation and covariance between X and residual values. Tips: Analyse…..Correlation…Bivariate…Option…}

P>0.05, suggesting NO significant autocorrelation between consecutive forecasting errors

(f)    U~N(0,σ2): The errors are **normally distributed** {Draw Histogram for the Unstandardized residuals. Do tests of Normality (Descriptive statistics…explore…dependent list: unstandardized residuals….Plots…normality plots with tests…).if p-values are high/insignificant, we CAN`T reject H0: Normality. This means the errors are normally distributed}

The errors from model2 are approximately normally distributed according to the histogram and the Shapiro normal test(p>0.05 => cannot refute null hypothesis that the residuals’ distribution is normal)

(g)  Is there any **multicollinearity problem:?**{(i) Look at the correlation matrix to confirm that there is no linear relationship between independent variables (ii) Check tolerance and VIF: Analyse ….Regression…linear….Statistics (check collinearity diagnostics). The tolerance should be above 0.10 and VIF should be below 10}

NO multicollinearity problem according to the diagnostics plot

Code:

library(quantmod):

library(timeSeries)

library(tseries)

library(forecast)

library(xts)

#loading data

TME\_STOCKPRICE = read.csv('~/Downloads/TME1.csv')

str(TME\_STOCKPRICE)

head(TME\_STOCKPRICE)

tail(TME\_STOCKPRICE)

View(TME\_STOCKPRICE)

length(TME\_STOCKPRICE$Date)

length(TME\_STOCKPRICE$Adj.Close)

## plot the adjusted stock price

library(ggplot2)

qplot(Date, Adj.Close,data = TME\_STOCKPRICE)

# Create data frame with date and adj.close price

TME\_adj\_prices <- TME\_STOCKPRICE[, "Adj.Close", drop = F]

# Create return and output it

n <- nrow(TME\_adj\_prices)

#TME\_ret <- ((TME\_adj\_prices[2:n, 1]-TME\_adj\_prices[1:n-1, 1])

# /TME\_adj\_prices[1:n-1, 1]) \* 100

TME\_ret <- diff(log(TME\_STOCKPRICE$Adj.Close))

#check summary statistics

library(moments)

skewness(TME\_ret)

kurtosis(TME\_ret)

#plot return data and price data

#plot(TME\_STOCKPRICE['Date'][2:n,1], TME\_ret, type="o", col="blue", lwd=2)

plot(TME\_STOCKPRICE$Adj.Close, type='l', ylab="TME\_adjusted\_close\_price",col="blue",lwd=2)

plot(TME\_ret, type="l", col="blue", lwd=2)

#rownames(TME\_adj\_prices) <-TME\_adj\_prices$Date

qplot(Date, Adj.Close,data = TME\_STOCKPRICE)

qplot(Date, log(Adj.Close),data = TME\_STOCKPRICE)

#plot ACF and PACF curves

acf(TME\_STOCKPRICE$Adj.Close, lag.max = 24)

pacf(TME\_STOCKPRICE$Adj.Close, lag.max = 24)

acf(TME\_ret\_train, lag.max = 24)

pacf(TME\_ret\_train, lag.max = 24)

#from inspection we should use AR model

### 2.AR Model ###

### test autocorrelation between lag return and return ###

#create 1 day lag return by AR model

TME\_ret\_train <- TME\_ret[0:200]

TME\_ret\_test <- TME\_ret[200:250]

arimaModel\_1 = arima(TME\_ret\_train, order = c(1, 0, 0))

#plot residuals and evaluate model

plot(arimaModel\_1$residuals)

Box.test(arimaModel\_1$residuals, lag = 1, type="Ljung-Box")

#predict from the model

prediction <- forecast::forecast(arimaModel\_1, h=50)

plot(prediction, type='l', col='blue', lwd=2)

lines(ts(TME\_ret\_train))

###3 Week of the Day Effect###

TME\_model2 <- arima(TME\_ret\_train, order = c(1, 0, 18), include.mean = FALSE, seasonal=list(order = c(1, 0, 18), period = 2))

prediction2 <- forecast::forecast(TME\_model2, h=255-201)

plot(prediction2, type='l', col='blue', lwd=2)

lines(ts(TME\_ret))

#check whether data shows seasonal effect

TME\_timeseries <- ts(TME\_STOCKPRICE$Adj.Close, frequency = 12)

TME\_timeseriescomponents <- decompose(TME\_timeseries)

plot(TME\_timeseriescomponents)

#test model: goodness of fit

Box.test(TME\_model2$residuals, lag = 5, type="Ljung-Box")

#test covariance between errors

summary(TME\_model2)

arch.test(TME\_model2, output=TRUE)

#test if variance of errors is cons

arch.test(TME\_model2, output=TRUE)

#test covariance between error and return

Box.test(TME\_model2$residuals, lag = 5, type="Ljung-Box")

#draw histogram for errors

hist(TME\_model2$residuals)

shapiro.test(TME\_model2$residuals) #p-value=0.3722 -> normal

#check multicolinearity

library(aTSA)

ts.diag(TME\_model2)