STAT 443: Forecasting, Fall 2022 Group Project

Predicting Real Estate Stock Prices in the United States (approximate one year, 240 working days)

2022-12-01

Jiacheng Zhang (20883204) - Introduction, Conclusion, PowerPoint Slides

Ping Qin (20796784) - Descriptive analysis of data, Smoothing models

Minghao Zhai (20825248) - Descriptive analysis of data, Box-Jenkins Models

Jodi Yuen Kei Mak (20857335) - Regression models, PowerPoint Slides

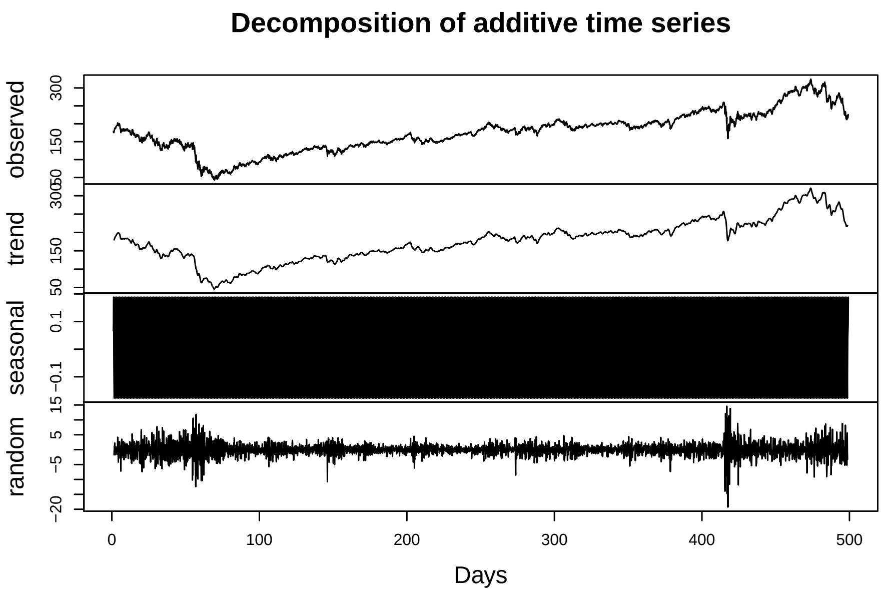
`(Title changed and we emailed to Prof Ramezan about it)`

Introduction and Motivation

We want to do the prediction for approximate one year, 240 working days, for Real Estate stock prices in the United States. Since stock prices of Real Estate indicate change of residence price which is people most concerned. Then people can use our project as reference, to decide when they buy their residence.

Data

We got our data from Yahoo finance, which records all price changes of stocks. We choose the adjusted price of Real Estate stock, from Jan 2007 to Oct 2022, as our data.

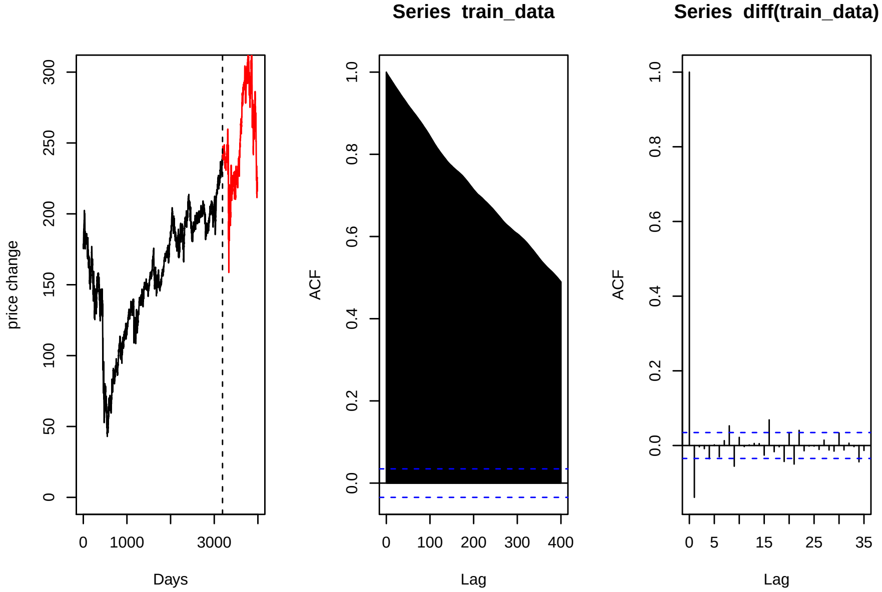


First Step: Data

We want to find if the data is stationary. Based on decomposition, we observed that trend component is significant because the range is wider than random component, but seasonal component is not significant. Therefore we can make a temporary assumption that trend is important component, but seasonal is not. From above, the data is not stationary.

Second Step: Polynomial Regression

We start doing a linear regression by we divide the data into two parts.(Train, Test). By the calculating the MSE, we can find the optimal regression model.



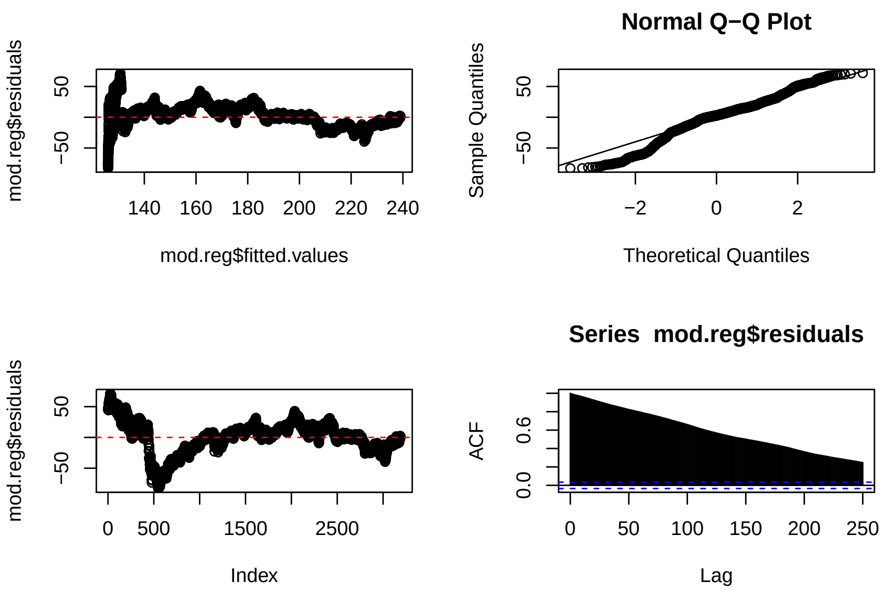
By the ACF of the data after differencing the trend, we observed that the period is 8.

```
##            mse.reg mse.reg.withseasonal
##  [1,] 1.623793e+03         1.623766e+03
##  [2,] 1.247046e+03         1.247078e+03
##  [3,] 1.925384e+04         1.925359e+04
##  [4,] 1.748213e+04         1.748304e+04
##  [5,] 1.012509e+04         1.012669e+04
##  [6,] 1.676010e+03         1.677329e+03
##  [7,] 4.096008e+06         4.096350e+06
##  [8,] 3.885899e+05         3.884622e+05
##  [9,] 1.338145e+06         1.336985e+06
## [10,] 1.724289e+08         1.724579e+08
```

```
## [1] 2
```

From above, The regression model without seasonal(degree = 2) has the lowest mse.

After finding the degree, we can do the linear regression.



For the first plot, we can see that the line fluctuated around 0.

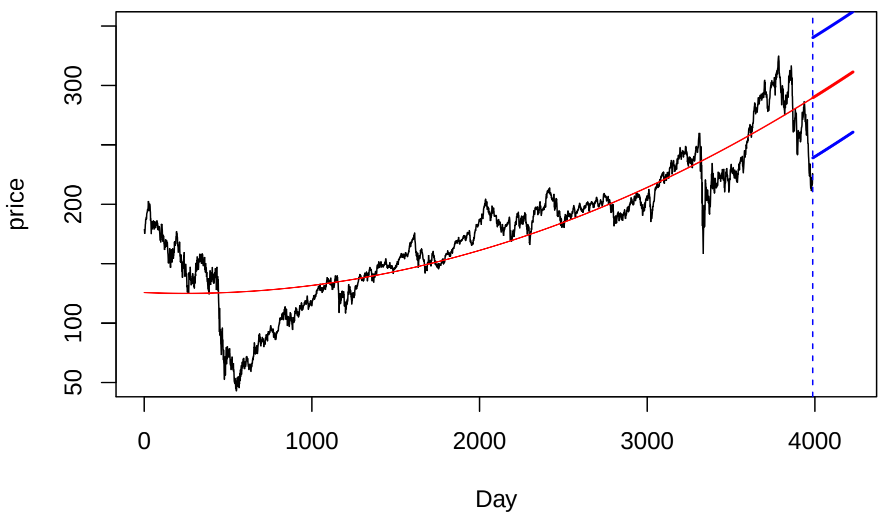
For the second plot, the line is almost a straight line, the normality is satisfactory.

For the third plot, it is obvious not constant variation. However, if we ignore the shortfall, we can see that after the shortfall the variation is around zero.

For the fourth plot, there is a decay which mean trend exist.

Hence we conclude that the linear regression may not be the best model.

Based on the regression, we predict the 240 working day of future stock price.



Third Step: Smoothing Methods

we want to transform the non-stationary data into stationary data using simple exponential smoothing, double exponential smoothing, additive smoothing, and multiplicative smoothing.

Mse of simple exponential smoothing

```
## [1] 1226.446
```

Mse of double exponential smoothing model

```
## [1] 1592.035
```

Mse of additive-HoltWinters model

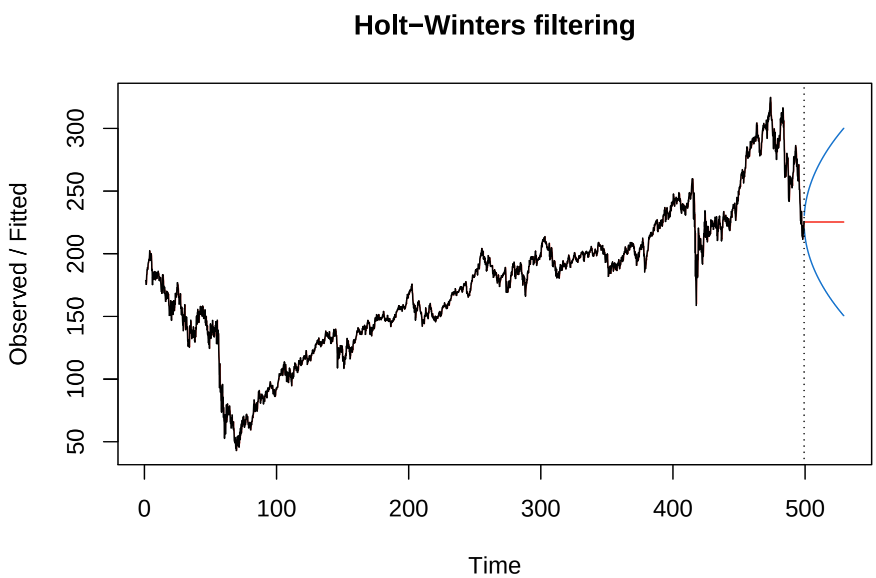
```
## [1] 5533.431
```

Mse of multiplicative-HoltWinters model

```
## [1] 5549.233
```

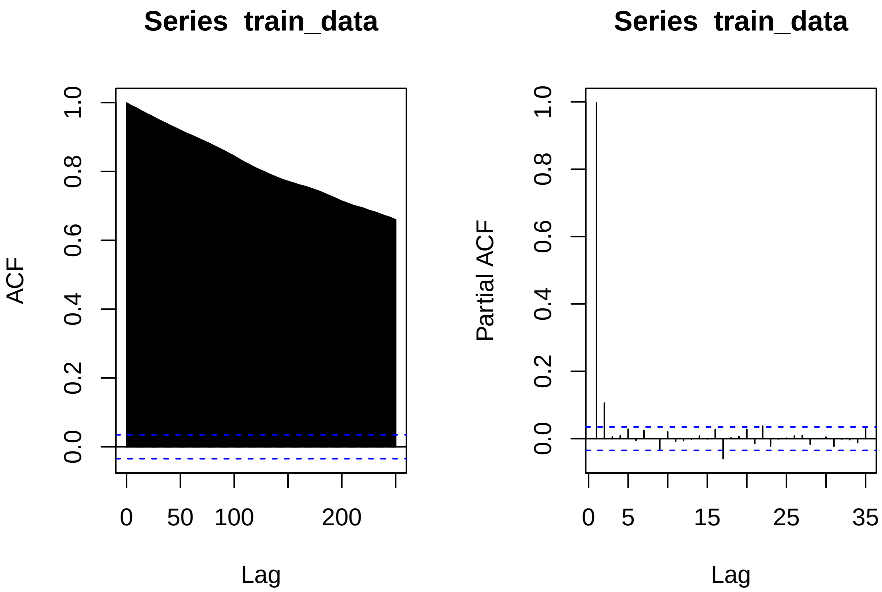
From above the mse of simple exponential smoothing is the lowest, so we choose simple exponential smoothing.

Then we fit a new model based on the new data, and predict the 240 working days of future stock price.

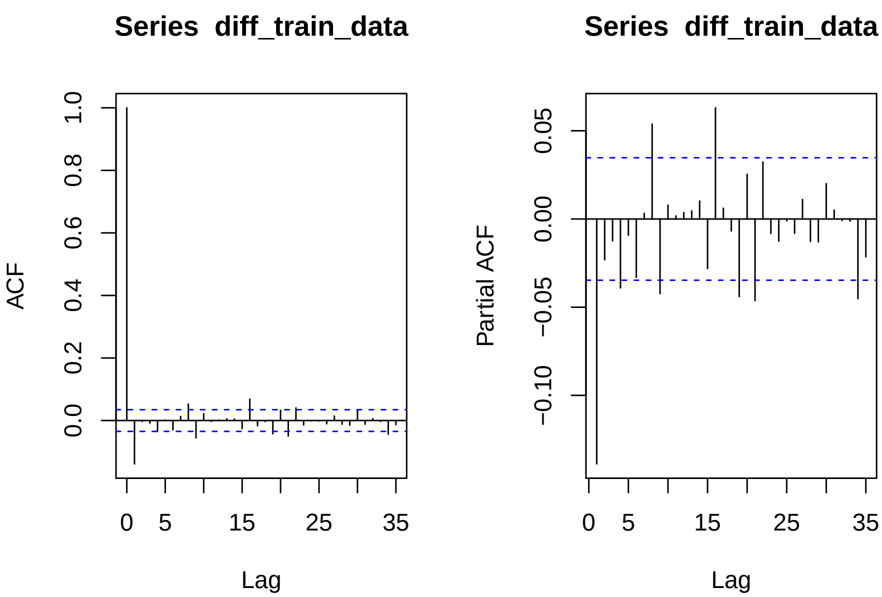


Forth Step: Box-Jenkins

we start to use Box-Jenkins method.



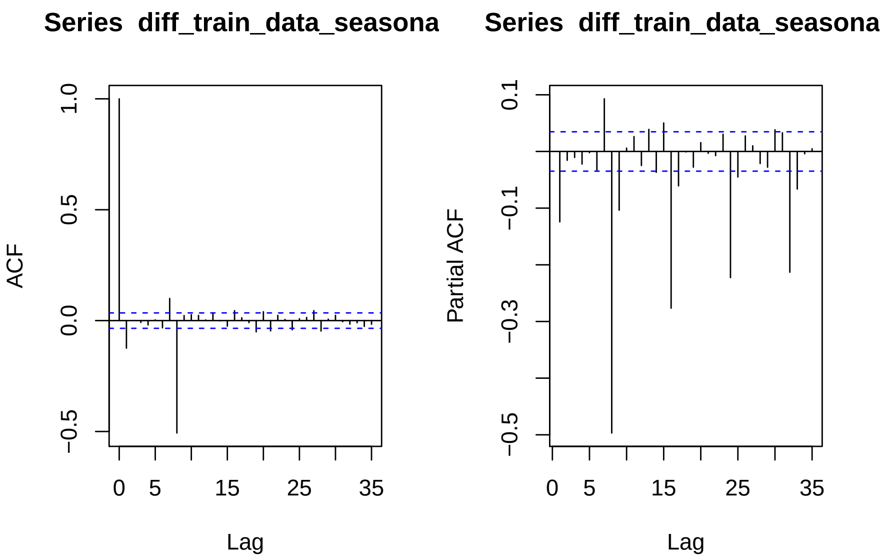
We plot the acf and pacf plot for the original data.



Then we apply differencng, and plot the acf and pacf for the data after differencing.

From the acf plot, We can see that trend is removed.

Then we want to remove the seasonality.

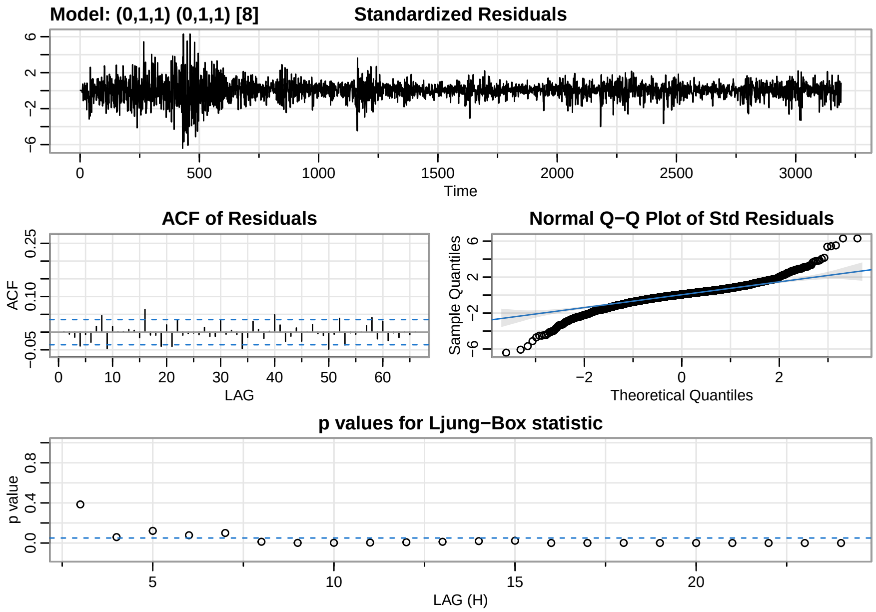


This is the acf and pacf plot after we remove the seasonality.

From observation s = 8.
From above, we remove trend one time and remove seasonality one time, so d = 1, D = 1.
Ignoring the non-seasonal lag, the acf cuts off after the first seasonal lag, and the pacf has an exponential decay, hence P = 0, Q = 1.
Ignoring the seasonal lags, the acf seems to cut off at lag 1 and pacf has damped sine wave, hence p=0, q =1.

From above, We propose a model(p=0,d=1,q=1,P=0,D=1,Q=1,S=8).

```
## initial  value 1.182797
## iter   2 value 0.974117
## iter   3 value 0.929455
## iter   4 value 0.885639
## iter   5 value 0.875176
## iter   6 value 0.875135
## iter   7 value 0.875123
## iter   8 value 0.875121
## iter   8 value 0.875121
## iter   8 value 0.875121
## final  value 0.875121
## converged
## initial  value 0.868516
## iter   2 value 0.863660
## iter   3 value 0.861285
## iter   4 value 0.861281
## iter   5 value 0.861196
## iter   6 value 0.861195
## iter   6 value 0.861195
## final  value 0.861195
## converged
```



For the first plot(standardized residuals), standardized residuals have zero , but not have constant variance.

For the second plot(ACF of Residuals), there is no significant lags, so the residuals are uncorrelative.

For the third plot(QQ plot), the points are not located on a straight line, so it not follows a normal distribution.

For the fourth plot(p value for Ljung-Box statistic), most of the points are under the blue lines, which means the model we choose is not good enough.

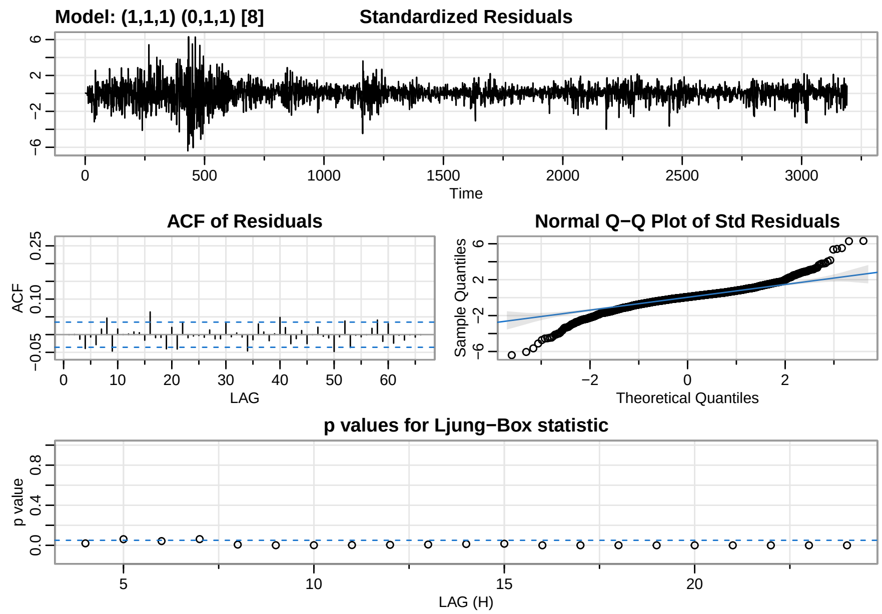
Then, we propose the second model.

Ignoring the seasonal lags, both the acf and the pacf show damped since wave pattern, hen p = 1, and q = 1.

From above, we propose another model(p=1,d=1,q=1,P=0,D=1,Q=1,S=8).

```
## initial  value 1.182952
## iter   2 value 0.981094
## iter   3 value 0.923590
## iter   4 value 0.883302
## iter   5 value 0.876657
## iter   6 value 0.875529
## iter   7 value 0.875479
## iter   8 value 0.875441
## iter   9 value 0.875440
## iter  10 value 0.875438
## iter  11 value 0.875435
## iter  12 value 0.875404
## iter  13 value 0.875377
## iter  14 value 0.875358
## iter  15 value 0.875354
## iter  16 value 0.875353
## iter  16 value 0.875353
## iter  16 value 0.875353
## final  value 0.875353
## converged
## initial  value 0.868522
## iter   2 value 0.863634
## iter   3 value 0.861331
```

```
## iter   4 value 0.861223
## iter   5 value 0.861165
## iter   6 value 0.861160
## iter   7 value 0.861159
## iter   8 value 0.861158
## iter   8 value 0.861158
## final  value 0.861158
## converged
```



For the first plot(standardized residuals), standardized residuals have zero , but not have constant variance.

For the second plot(ACF of Residuals), there is no significant lags, so the residuals are uncorrelative.

For the third plot(QQ plot), the points are not located on a straight line, so it not follows a normal distribution.

For the fourth plot(p value for Ljung-Box statistic), most of the points are under the blue lines, which means the model we choose is not good enough.

Compare the AIC, AICC ,and BIC, we choose a better model.

```r
c(model1$AIC,model1$AICc,model1$BIC)
```

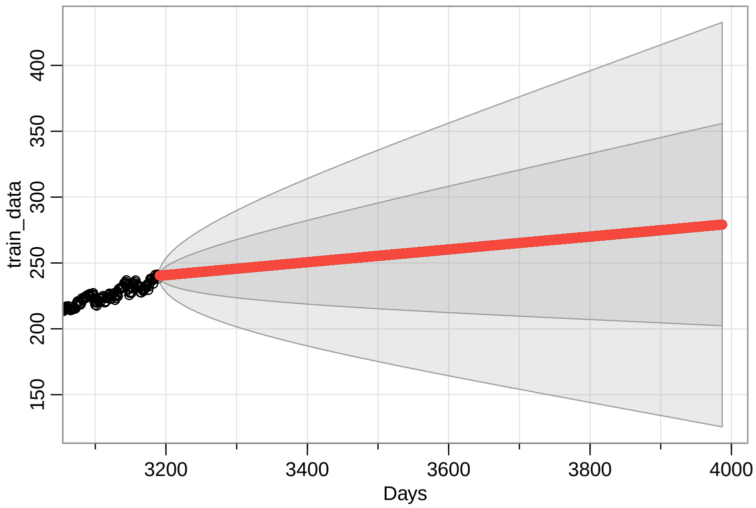
```
## [1] 4.562154 4.562155 4.567874
```

```r
c(model2$AIC,model2$AICc,model2$BIC)
```

```
## [1] 4.562709 4.562711 4.570335
```

From above, model1(p=1,d=1,q=1,P=0,D=1,Q=1,S=8) has the lower value.

Calculate the Mse of model1 to compare with the mse of other models we found before.

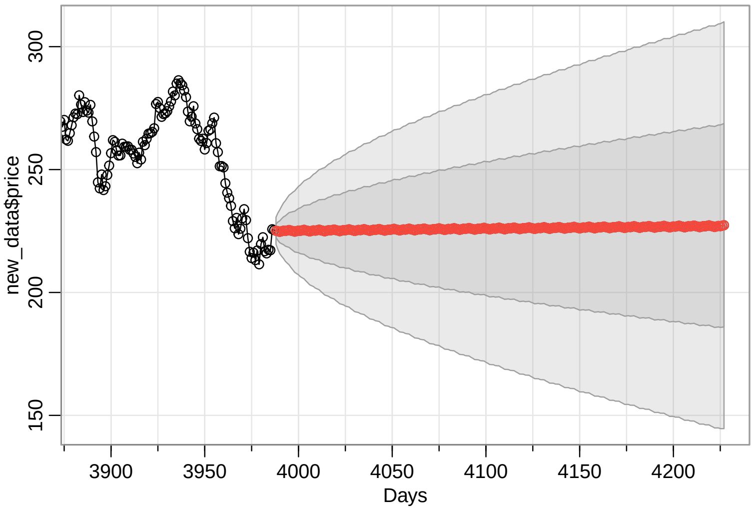



```
## [1] 784.9541
```

The mse of the linear regression (degree =2) is 1.247046e+03.
The mse of the simple exponential smoothing model is1226.446.
The mse of the Box-Jenkins model is 784.9541.
Now we have three different MSE for various methods. By observation, we know that Box-jenkins has the smallest MSE in general. Therefore we pick Box-jenkins as the final model.

Final Step: Prediction based on the best model chosen

Using Box-Jenkins to predict the 240 working day of future stock price.




```
## $pred
## Time Series:
## Start = 3988
## End = 4227
## Frequency = 1
##   [1] 225.0870 224.9279 224.6848 224.9013 225.1131 225.0184 225.2087 225.4080
##   [9] 225.1543 224.9952 224.7521 224.9686 225.1804 225.0857 225.2760 225.4753
##  [17] 225.2216 225.0625 224.8194 225.0359 225.2477 225.1530 225.3433 225.5426
##  [25] 225.2889 225.1298 224.8867 225.1032 225.3150 225.2203 225.4106 225.6099
##  [33] 225.3562 225.1971 224.9540 225.1705 225.3823 225.2876 225.4779 225.6772
##  [41] 225.4235 225.2644 225.0213 225.2378 225.4496 225.3549 225.5452 225.7445
##  [49] 225.4908 225.3317 225.0886 225.3051 225.5169 225.4222 225.6125 225.8118
##  [57] 225.5581 225.3990 225.1559 225.3724 225.5842 225.4895 225.6798 225.8791
##  [65] 225.6254 225.4663 225.2232 225.4397 225.6515 225.5568 225.7471 225.9464
##  [73] 225.6927 225.5336 225.2905 225.5070 225.7188 225.6241 225.8144 226.0137
##  [81] 225.7599 225.6009 225.3578 225.5743 225.7861 225.6914 225.8817 226.0810
##  [89] 225.8272 225.6682 225.4251 225.6416 225.8534 225.7587 225.9490 226.1483
##  [97] 225.8945 225.7355 225.4924 225.7089 225.9207 225.8260 226.0163 226.2156
## [105] 225.9618 225.8028 225.5597 225.7762 225.9880 225.8933 226.0836 226.2829
## [113] 226.0291 225.8701 225.6270 225.8435 226.0553 225.9606 226.1509 226.3502
## [121] 226.0964 225.9374 225.6943 225.9108 226.1226 226.0279 226.2182 226.4175
## [129] 226.1637 226.0047 225.7616 225.9781 226.1899 226.0952 226.2855 226.4848
## [137] 226.2310 226.0720 225.8289 226.0454 226.2572 226.1625 226.3528 226.5521
## [145] 226.2983 226.1393 225.8962 226.1127 226.3245 226.2298 226.4201 226.6194
## [153] 226.3656 226.2066 225.9635 226.1800 226.3918 226.2971 226.4874 226.6867
## [161] 226.4329 226.2739 226.0308 226.2473 226.4591 226.3644 226.5547 226.7540
## [169] 226.5002 226.3412 226.0981 226.3146 226.5264 226.4317 226.6220 226.8213
## [177] 226.5675 226.4085 226.1654 226.3819 226.5937 226.4990 226.6893 226.8886
## [185] 226.6348 226.4758 226.2327 226.4492 226.6610 226.5663 226.7566 226.9559
## [193] 226.7021 226.5431 226.3000 226.5165 226.7283 226.6336 226.8239 227.0232
## [201] 226.7694 226.6104 226.3673 226.5838 226.7956 226.7009 226.8912 227.0905
## [209] 226.8367 226.6777 226.4346 226.6511 226.8629 226.7682 226.9585 227.1578
## [217] 226.9040 226.7450 226.5019 226.7184 226.9302 226.8355 227.0258 227.2251
## [225] 226.9713 226.8123 226.5692 226.7857 226.9975 226.9028 227.0931 227.2924
```

```
## [233] 227.0386 226.8796 226.6365 226.8530 227.0648 226.9701 227.1604 227.3597
##
## $se
## Time Series:
## Start = 3988
## End = 4227
## Frequency = 1
##   [1]  2.763224  3.702833  4.448206  5.085480  5.651342  6.165488  6.639940
##   [8]  7.082680  7.504475  7.903263  8.282873  8.645831  8.994155  9.329482
##  [15]  9.653168  9.966346 10.273759 10.571839 10.861741 11.144105 11.419489
##  [22] 11.688387 11.951236 12.208426 12.463443 12.713012 12.957775 13.198001
##  [29] 13.433931 13.665788 13.893776 14.118083 14.341619 14.561433 14.777977
##  [36] 14.991393 15.201814 15.409361 15.614149 15.816286 16.018337 16.217607
##  [43] 16.414458 16.608976 16.801242 16.991333 17.179320 17.365271 17.551514
##  [50] 17.735560 17.917715 18.098036 18.276579 18.453394 18.628531 18.802035
##  [57] 18.976057 19.148272 19.318951 19.488136 19.655865 19.822174 19.987099
##  [64] 20.150674 20.314907 20.477610 20.639032 20.799200 20.958145 21.115893
##  [71] 21.272470 21.427903 21.584086 21.738946 21.892711 22.045403 22.197045
##  [78] 22.347658 22.497262 22.645877 22.795303 22.943564 23.090873 23.237249
##  [85] 23.382708 23.527268 23.670944 23.813753 23.957413 24.100034 24.241816
##  [92] 24.382773 24.522920 24.662270 24.800836 24.938633 25.077307 25.215042
##  [99] 25.352030 25.488281 25.623807 25.758621 25.892731 26.026150 26.160464
## [106] 26.293923 26.426707 26.558828 26.690294 26.821117 26.951302 27.080862
## [113] 27.211328 27.341008 27.470075 27.598538 27.726406 27.853688 27.980389
## [120] 28.106518 28.233560 28.359873 28.485627 28.610827 28.735482 28.859599
## [127] 28.983183 29.106242 29.230215 29.353511 29.476290 29.598560 29.720328
## [134] 29.841598 29.962376 30.082670 30.203877 30.324450 30.444545 30.564169
## [141] 30.683326 30.802022 30.920261 31.038049 31.156750 31.274854 31.392514
## [148] 31.509734 31.626520 31.742877 31.858806 31.974316 32.090734 32.206589
## [155] 32.322029 32.437058 32.551680 32.665900 32.779721 32.893147 33.007479
## [162] 33.121276 33.234684 33.347706 33.460346 33.572609 33.684495 33.796011
## [169] 33.908428 34.020336 34.131877 34.243055 34.353873 34.464335 34.574442
## [176] 34.684199 34.794851 34.905019 35.014840 35.124317 35.233455 35.342255
## [183] 35.450720 35.558854 35.667876 35.776435 35.884666 35.992571 36.100154
## [190] 36.207417 36.314361 36.420991 36.528504 36.635573 36.742329 36.848777
## [197] 36.954917 37.060754 37.166288 37.271522 37.377633 37.483317 37.588704
## [204] 37.693796 37.798596 37.903107 38.007327 38.111263 38.216069 38.320464
## [211] 38.424575 38.528405 38.631956 38.735230 38.838227 38.940952 39.044541
## [218] 39.147732 39.250653 39.353304 39.455689 39.557808 39.659663 39.761256
## [225] 39.863707 39.965774 40.067582 40.169131 40.270424 40.371464 40.472248
## [232] 40.572783 40.674168 40.775182 40.875947 40.976463 41.076734 41.176761
## [239] 41.276543 41.376085
```

Conclusion

Linear regression model, simple exponential smoothing model, and Box-Jenkins model do not fit the data well since they have their own limitations. However based on the mse of the three models, we choose the Box-Jenkins as our final model to predict the price.

Conclusion in context

Based on we best model we choose, the stock price of Real Estate is non-periodic. Then, with the prediction we do, the future stock price will stable around 225 in next 240 working days. The lowest stock price will be around 190, and the highest stock price will be around 270 with 95% probability.