#

# In-class Exercise: Fitting Dow Jones Data with ARIMA

#

#########################################################

# Use This file as a template for your assignment.

# Submitt your code and comments together with (selected) output from R console.

# No need to submit the plots.

# Your Name: \_\_\_\_\_\_\_Nao Mimoto\_\_\_\_\_\_\_\_\_\_\_\_\_

# 1.

# -----------------

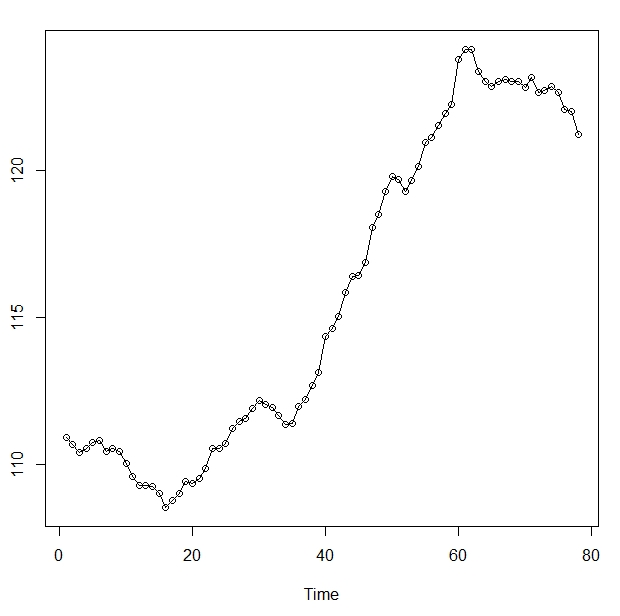
# Read in dowj.txt file. Plot the time series.

# Does the plot looks stationary?

# Plot ACF and PACF of the series.

# Test for the stationarity using Augmented Dickey-Fuller Unit-Root test.

D <- read.csv("http://gozips.uakron.edu/~nmimoto/pages/datasets/dowj.csv")



# 2.

# -----------------

# Take the difference of dowj data. Plot the time series.

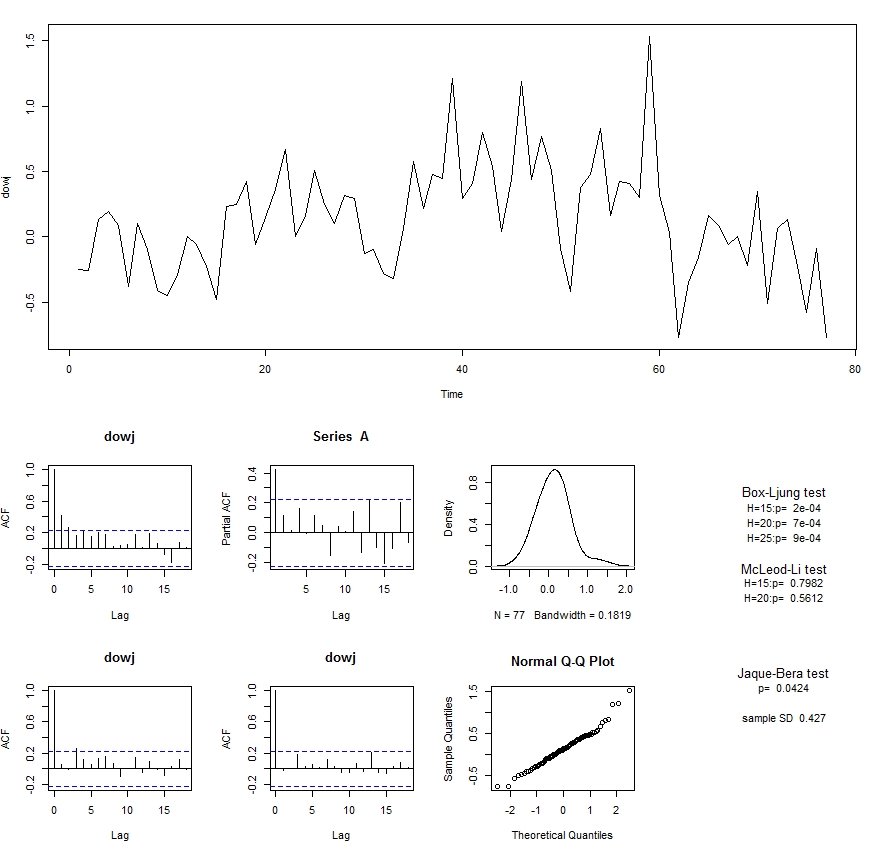
# Does the plot looks stationary?

# Plot ACF and PACF of the series.

# What does ADF test say about stationarity?

D2 <- ts(diff(D1), start=c(1,1), freq=1)

Randomness.tests(D2)



adf.test(D2)

Augmented Dickey-Fuller Test

data: D2

Dickey-Fuller = -2.0339, Lag order = 4, p-value = 0.5617

alternative hypothesis: stationary

pp.test(D2)

Phillips-Perron Unit Root Test

data: D2

Dickey-Fuller Z(alpha) = -41.3714, Truncation lag parameter = 3,

p-value = 0.01

alternative hypothesis: stationary

kpss.test(D2)

KPSS Test for Level Stationarity

data: D2

KPSS Level = 0.3486, Truncation lag parameter = 2, p-value = 0.09932

ADF test does not have enough evidence to reject the null of Non-stationarity.

PP test has rejected the null of non-stationarity.

KPSS test cannot reject the null of stationarity.

ADF test is supporting the non-stationarity, but both PP test and KPSS tests are supporting stationarity.

# 3.

# -----------------

# Take additional difference of dowj data. Plot the time series.

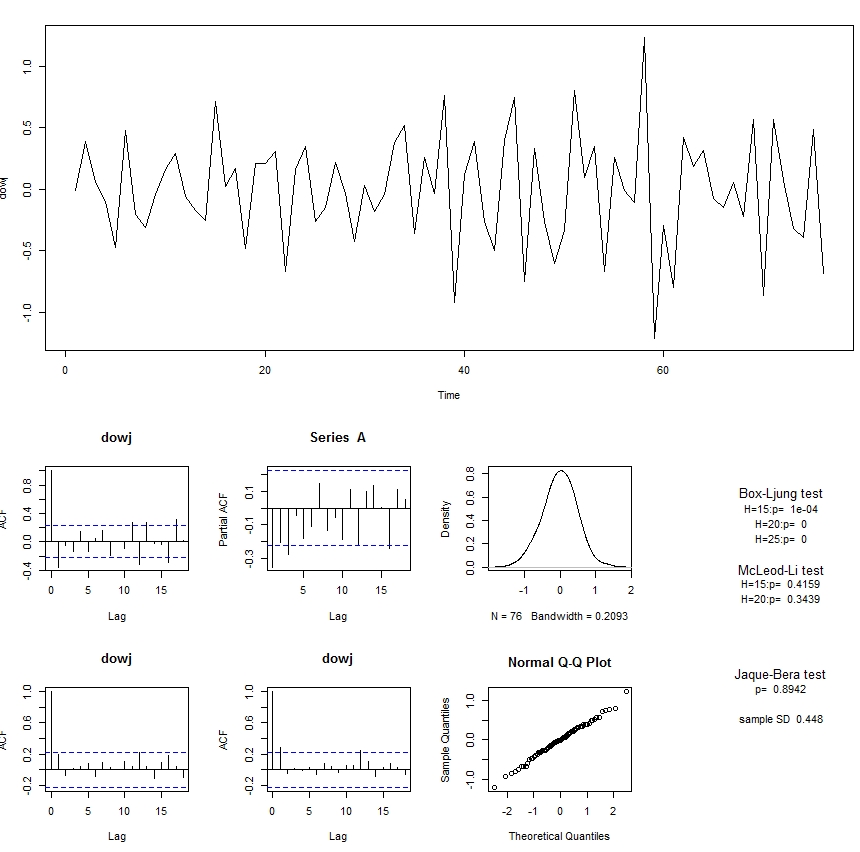
# Does the plot looks stationary?

# Plot ACF and PACF of the series.

# What does ADF test say about stationarity?

D3 <- ts(diff(D2), start=c(1,1), freq=1)

Randomness.tests(D3)



adf.test(D3)

Augmented Dickey-Fuller Test

data: D3

Dickey-Fuller = -5.9052, Lag order = 4, p-value = 0.01

alternative hypothesis: stationary

pp.test(D3)

Phillips-Perron Unit Root Test

data: D3

Dickey-Fuller Z(alpha) = -86.7696, Truncation lag parameter = 3,

p-value = 0.01

alternative hypothesis: stationary

kpss.test(D3)

KPSS Test for Level Stationarity

data: D3

KPSS Level = 0.0854, Truncation lag parameter = 2, p-value = 0.1

ADF and PP test has rejected the null of non-stationarity.

KPSS could not reject the null of stationarity.

We conclude that the second difference is stationary, given the consistency of the three tests.

# 2-b.

# -----------------

# Now based on what we saw in question 2, model the origianl dowj data

# with ARIMA(p, 1, q).

# Use auto.arima() in forecast package to choose p and q based on AICc.

# Diagnose the model fit. Is the model fitting well? All the parameters significant?

# If not, manually search for better value of p and q.

Est1 <- auto.arima(D1, d=1, seasonal=FALSE); Est1

Series: D1

ARIMA(1,1,1)

Coefficients:

ar1 ma1

0.8510 -0.5263

s.e. 0.1383 0.2548

sigma^2 estimated as 0.1434: log likelihood=-34.69

AIC=75.38 AICc=75.71 BIC=82.41

Randomness.tests(Est1$residuals)

BL15 0.287

BL20 0.088

BL25 0.094

ML15 0.701

ML20 0.522

JB 0.048

SD 0.377

Auto.arima has chosen ARIMA(1,1,1) based on lowest AICC. However, parameter for MA1 is barely significant, and residuals are barely passing the Ljung-Box test at H=20 and 25 (they pass at size 5%, but fail at size 10%). Model fit is adequate but not great.

Model ARIMA(1,1,0) is tried, but could not pass L-B test at size 5%.

For d=1, ARIMA(1,1,1) is the best we have.

# 2-c.

# -----------------

# Using the model you came up in the previous question, give 5-day prediction

# of dowj value. Plot the data(black) and predictioin(red) on the same plot.

# The range of x-axis must be suitably chosen.

If we use ARIMA(1,1,1) above, I get prediction as:

est <- auto.arima(D1, d=1, seasonal=FALSE); est

x.p <- predict(est, n.ahead=5, se.fit=TRUE)

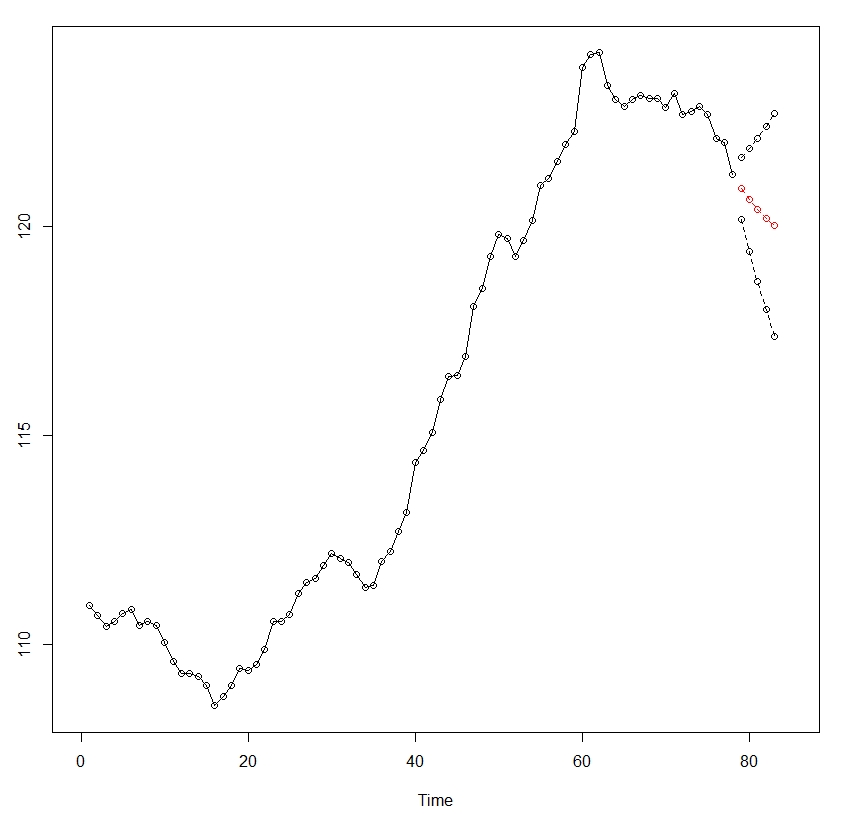
x.pred <- ts(x.p$pred, start=79)

x.se1 <- ts(x.pred + 1.96\*x.p$se, start=79)

x.se2 <- ts(x.pred - 1.96\*x.p$se, start=79)

ts.plot(cbind(D1, x.pred, x.se1, x.se2), type="o",

col=c('black','red', 'black', 'black'), lty=c(1,2,2,2), xlim=c(0,85))



# 2-d.

# -----------------

# In part (2-b), your ARIMA parameter estimation gave standard errors for

# estimation. Can you trust that number? Why? How would you verify?

The s.e. of the parameter estimate is calculated using asymptotic variance formula for MLE as n goes to infinity. With n=78, there’s not much reason to believe in the number printed on screen. To bet better estimate of the standard deviation of the parameter estimate, one can perform Monte Carlo simulation with estimated parameter. Here is example:

# 3-b.

# -----------------

# Now based on what we saw in question 3, model the original dowj data

# with ARIMA(p, 2, q).

# Use auto.arima() in forecast package to choose p and q based on AICc.

# Diagnose the model fit. Is the model fitting well? All the parameters significant?

# If not, manually search for better value of p and q.

Est1 <- auto.arima(D1, d=2, seasonal=FALSE); Est1

Series: D1 ARIMA(1,2,1)

ar1 ma1

0.2479 -0.8391

s.e. 0.1447 0.0819

sigma^2 estimated as 0.1449: log likelihood=-34.85

AIC=75.7 AICc=76.03 BIC=82.69

Randomness.tests(Est1$residuals)

BL15 0.278

BL20 0.069

BL25 0.065

ML15 0.605

ML20 0.436

JB 0.231

SD 0.379

Est1 <- arima(D1, order=c(0,2,1)); Est1

Series: D1

ARIMA(0,2,1)

Coefficients:

ma1

-0.7157

s.e. 0.1133

sigma^2 estimated as 0.1504: log likelihood=-36.2

AIC=76.4 AICc=76.57 BIC=81.06

Randomness.tests(Est1$residuals)

BL15 0.282

BL20 0.073

BL25 0.095

ML15 0.397

ML20 0.396

JB 0.152

SD 0.386

# 3-c.

# -----------------

# Using the model you came up in the previous question, give 5-day prediction

# of dowj value. Plot the data(black) and predictioin(red) on the same plot.

# The range of x-axis must be suitablly chosen.

est <- arima(D1, order=c(0,2,1) ); est

x.p2 <- predict(est, n.ahead=5, se.fit=TRUE)

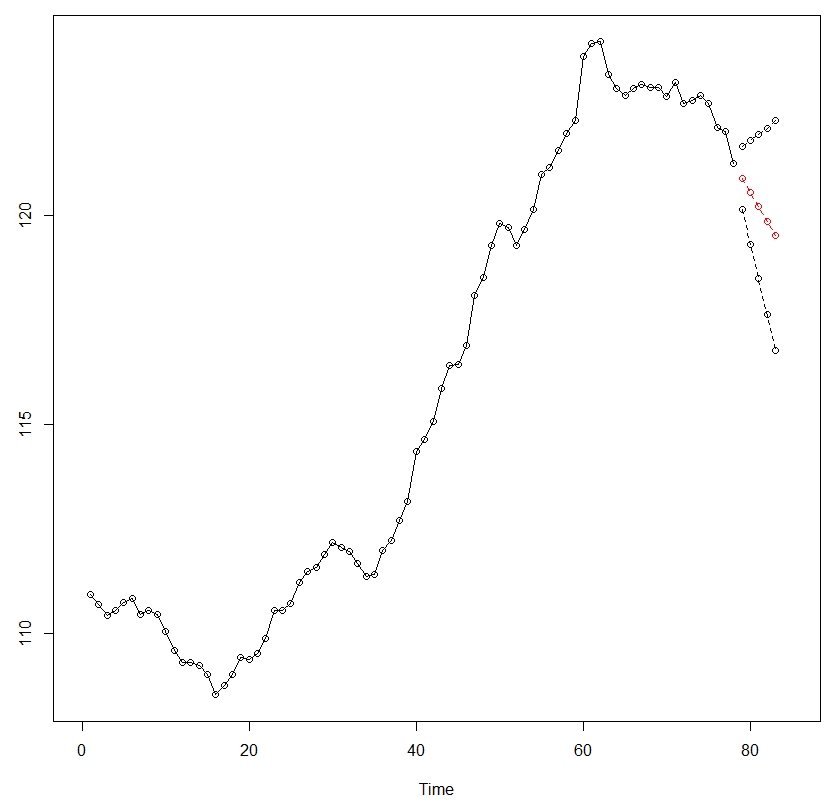
x.pred2 <- ts(x.p2$pred, start=79)

x.se12 <- ts(x.pred2 + 1.96\*x.p2$se, start=79)

x.se22 <- ts(x.pred2 - 1.96\*x.p2$se, start=79)

ts.plot(cbind(D1, x.pred2, x.se12, x.se22), type="o",

col=c('black','red', 'black', 'black'), lty=c(1,2,2,2), xlim=c(0,85))



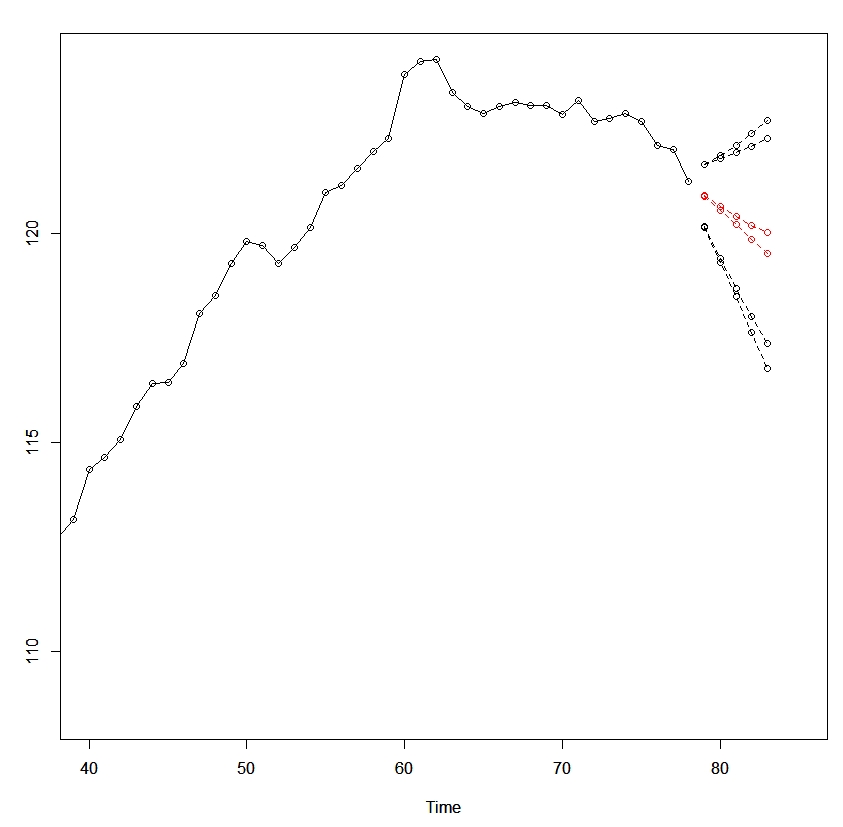
ts.plot(cbind(D1, x.pred, x.se1, x.se2), type="o",

col=c('black','red', 'black', 'black'), lty=c(1,2,2,2), xlim=c(40,85))

par(new=T)

ts.plot(cbind(D1, x.pred2, x.se12, x.se22), type="o",

col=c('black','red', 'black', 'black'), lty=c(1,2,2,2), xlim=c(40,85))



Difference between predictions form two models. Above is ARIMA(1,1,1).

# 4. (optional)

# -----------------

# Can you come up with some other way of fitting the dowj model?

#

Taking log-difference produces almost the same result as 2-b using auto.arima, with slightly improved number for L-B test.

Est1 <- auto.arima( log(D1) ); Est1

Series: log(D1)

ARIMA(1,1,1)

Coefficients:

ar1 ma1

0.8530 -0.5265

s.e. 0.1313 0.2424

sigma^2 estimated as 1.036e-05: log likelihood=332.42

AIC=-658.84 AICc=-658.51 BIC=-651.81

Randomness.tests(Est1$residuals)

BL15 0.356

BL20 0.101

BL25 0.104

ML15 0.767

ML20 0.569

JB 0.233

SD 0.003

# 5.

# -----------------

# Which model do you like better (2-b), (3-b) or 4?

# Why?

ARIMA(1,1,1) has one less difference, and smaller sigma hat. However, has problem with stationarity.

The p-values for L-B tests of ARIMA(0,2,1) is not any better than that of ARIMA(1,1,1). Therefore, between the two model, ARIMA(0,2,1) is the better model.

Note that you can not use AICC to choose between ARIMA(1,1,1) and ARIMA(0,2,1). You can use AICC to choose value of p and q for given d.