Supply Chain HW I2

Brett Scroggins

Due: 11/21/2017

```
library(fpp)
library(dplyr)
```

```
PG <- read.csv("/Users/brettscroggins/Downloads/IPG2211N.csv") %>%
  select(-DATE) %>%
  ts(start=c(1972,1), frequency=12)

PG1.tr <- window(PG, end=c(1995,12))
PG1.te <- window(PG, start=c(1996,1), end=c(2000,12))
```

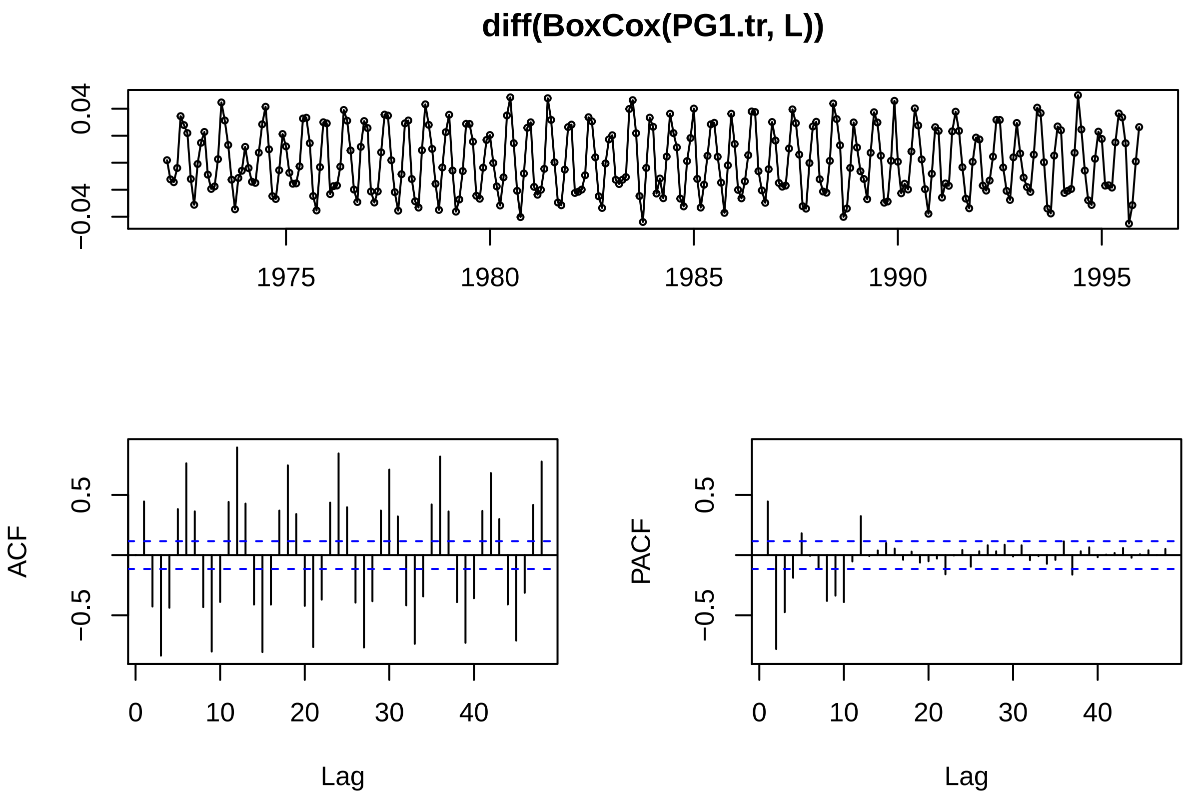
Question 1

```
# BoxCox Lambda for PG1
L = BoxCox.lambda(PG1.tr)
L
```

```
## [1] -0.2542538
```

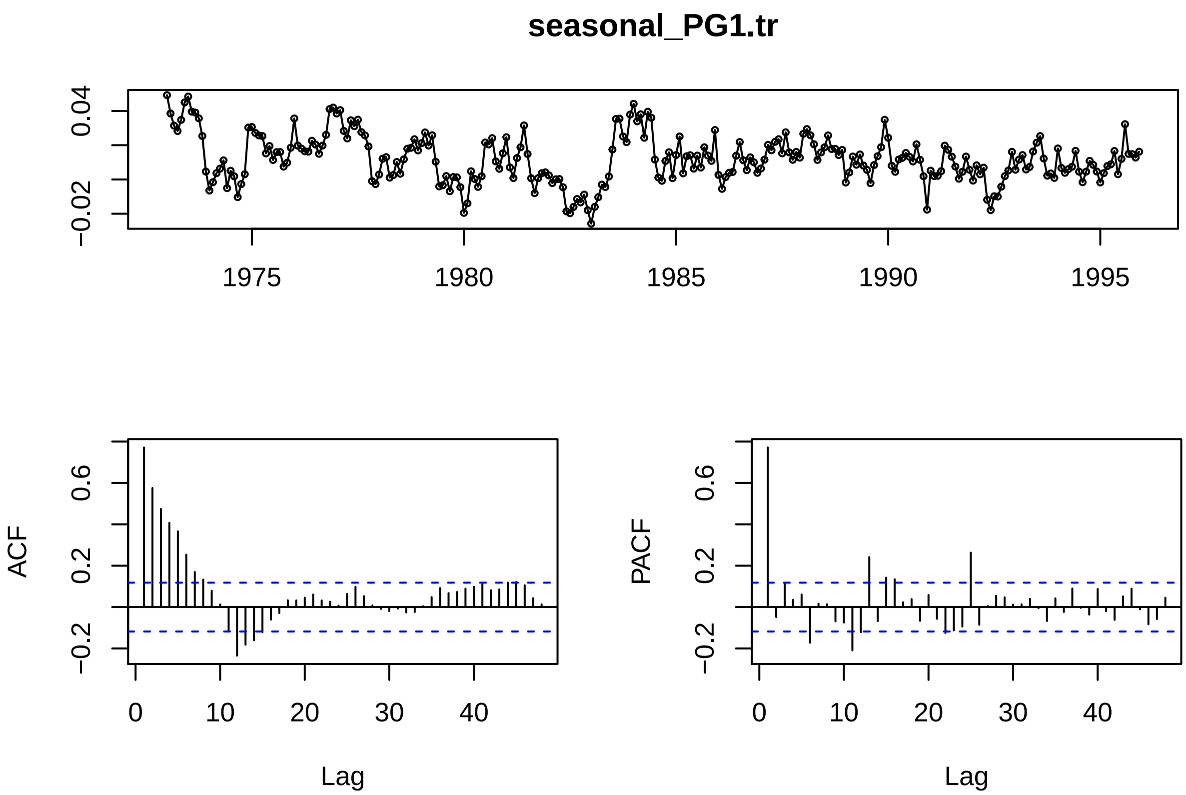
Part B

```
# d = 2
tsdisplay(diff(BoxCox(PG1.tr,L)), lag=48)
```



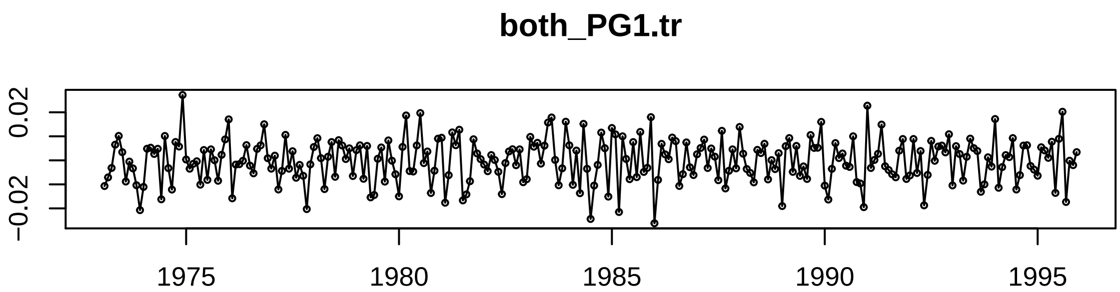
Part C

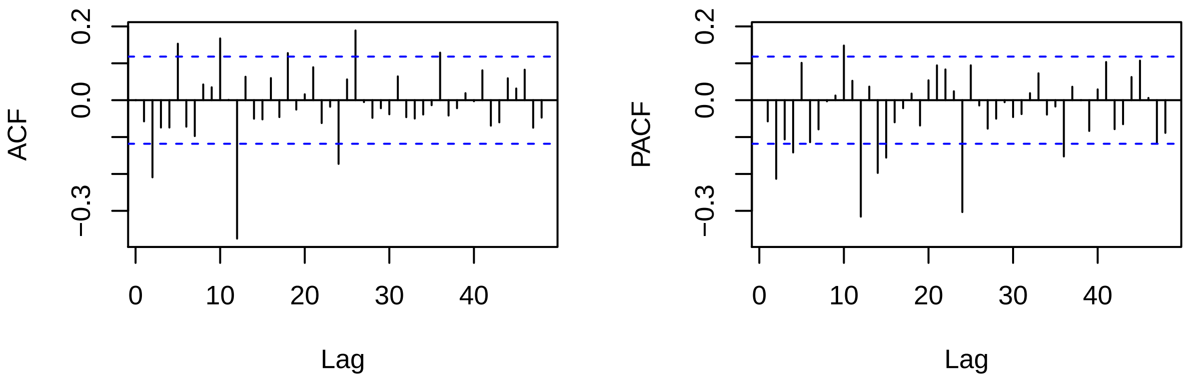
```
# D = 1
seasonal_PG1.tr = diff(BoxCox(PG1.tr,L),12)
tsdisplay(seasonal_PG1.tr, lag=48)
```



Part D

```r
# d = 1, D = 1
both_PG1.tr = diff(seasonal_PG1.tr)
tsdisplay(both_PG1.tr, lag=48)
```





Part E

```r
# adf test
adf.test(both_PG1.tr)
```

```
## Warning in adf.test(both_PG1.tr): p-value smaller than printed p-value
##
##  Augmented Dickey-Fuller Test
##
## data:  both_PG1.tr
## Dickey-Fuller = -7.9688, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary
```

What do you conclude from the test?

Solution

What can be seen from this test is that we must reject the null hypothesis that this is a stationary time series. The p-value returned is significant and therefore allows us to say this this is not a stationary series.

Part F

If you were to fit an ARIMA model to each of the (three) differenced series you obtained above, what would be the maximum order of the (p,d,q) (P,D,Q)_12 model in each case? (i.e., what is the maximum values of p,P,q and Q for each of the value combinations of d and D?)

Solution

- From the non-seasonally differenced model when d = 1: p = 5, q = 3, P = 1, Q = 4
- From the seasonally difference model when D = 1: p = 2, q = 7, P = 1, Q = 1
- From the model with both non-seasonal difference (d = 1) and seasonal difference (D = 1): p = 2, q = 2, P = 3, Q = 4

Question 2

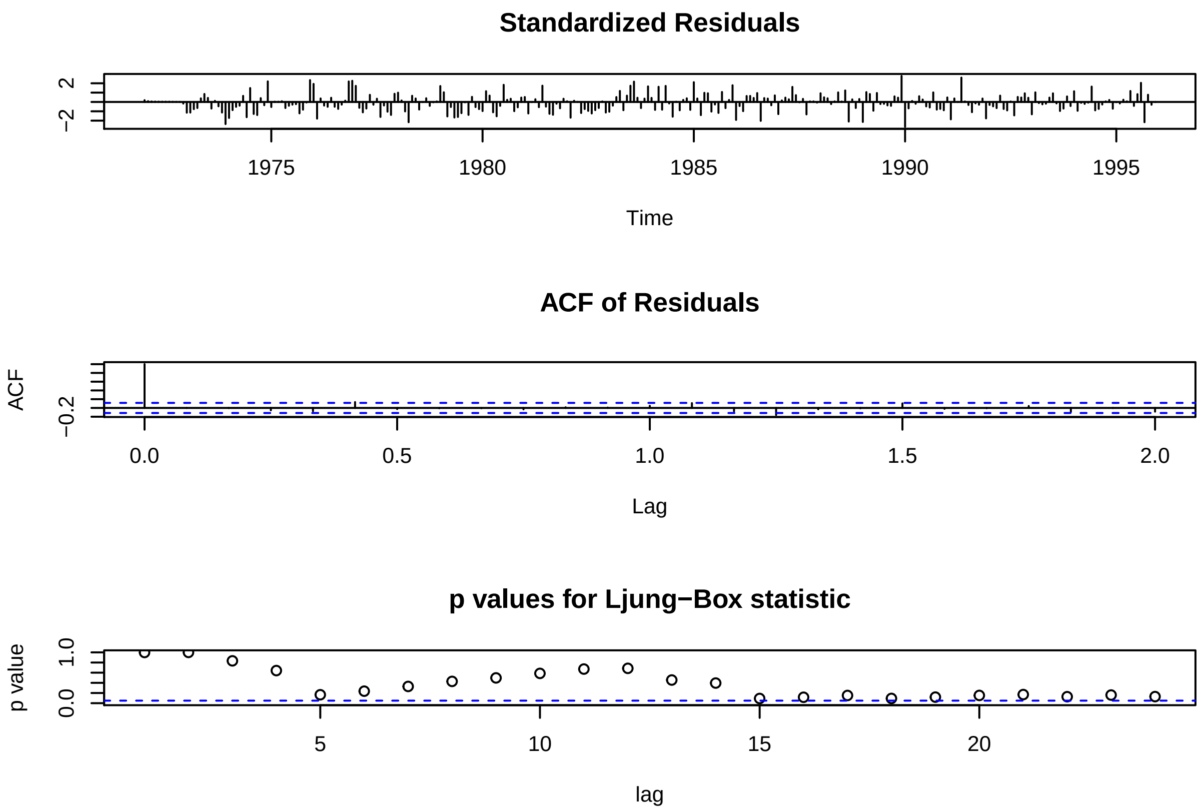
Part A

```
# auto arima
auto_arima = auto.arima(PG1.tr, lambda = L)
summary(auto_arima)
```

```
## Series: PG1.tr
## ARIMA(1,1,2)(0,1,1)[12]
## Box Cox transformation: lambda= -0.2542538
##
## Coefficients:
##          ar1      ma1      ma2     sma1
##       0.3733  -0.4466  -0.2637  -0.7941
## s.e.  0.1541   0.1522   0.0728   0.0476
##
## sigma^2 estimated as 4.947e-05:  log likelihood=969.56
## AIC=-1929.12   AICc=-1928.89   BIC=-1911.03
##
## Training set error measures:
##                       ME     RMSE       MAE        MPE    MAPE      MASE
## Training set -0.09697521 1.179734 0.8761734 -0.2013992 1.48562 0.3886194
##                     ACF1
## Training set -0.1091666
```

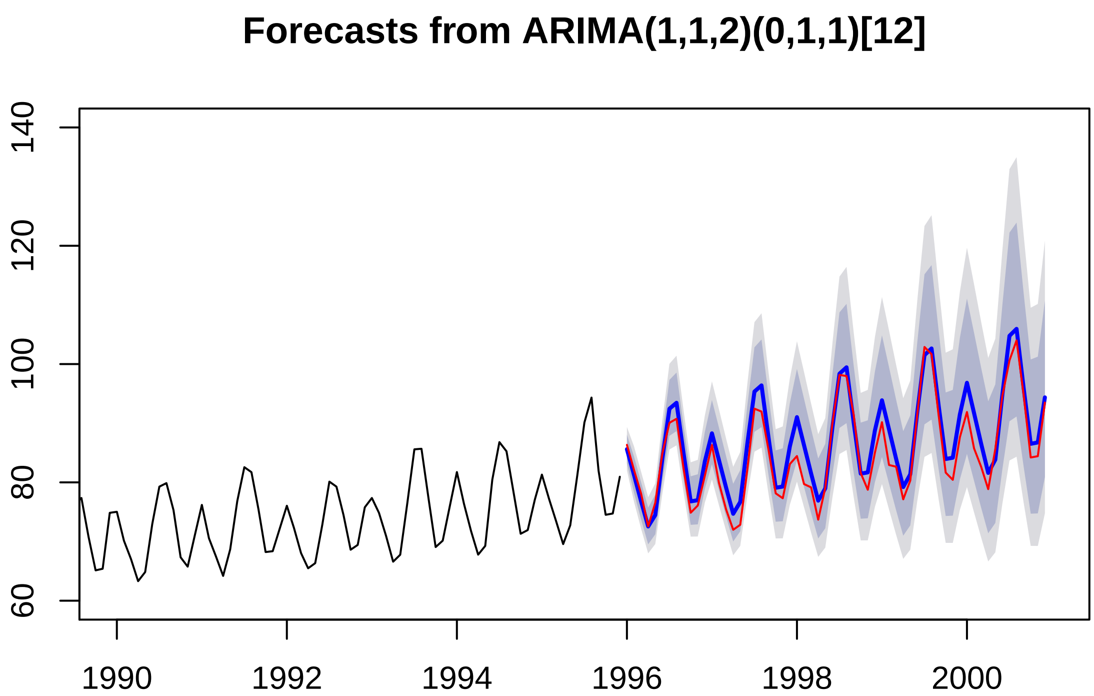
Part B

```
# tsdiag from Problem 1
tsdiag(auto_arima, gof.lag = 24)
```



Part C

```
# 60 month forecast
forecast_60months = forecast(auto_arima, h = 60)
plot(forecast_60months,xlim=c(1990,2001),ylim=c(60,140))
lines(PG1.te, col='red')
```



Part D

```r
# forecast accuracy
accuracy(forecast_60months, PG1.te)
```

```
##                       ME     RMSE       MAE        MPE     MAPE      MASE
## Training set -0.09697521 1.179734 0.8761734 -0.2013992 1.485620 0.3886194
## Test set     -2.03164696 2.895616 2.4081906 -2.4293395 2.872239 1.0681328
##                    ACF1 Theil's U
## Training set -0.1091666        NA
## Test set      0.5831711 0.4666031
```

Based on the visual inspection of the forecast plot and the out-of-sample fit statistics comment on the forecast bias.

Solution

Based on the graph, we can see that the prediction is slightly below the blue testing line. With this, and seeing that the ME grows significantly from training to testing set, it can be see that the projection is more bias towards smaller growth than what is actual.

Question 3

Part A

```r
# manual arima for seasonal
P3_arima = auto.arima(seasonal_PG1.tr, d=0, D=1, max.p = 2, max.q = 6, max.P = 1, max.Q = 1)
summary(P3_arima)
```

```
## Series: seasonal_PG1.tr
## ARIMA(1,0,0)(1,1,0)[12] with drift
##
## Coefficients:
##          ar1     sar1   drift
##       0.7472  -0.5863  -1e-04
## s.e.  0.0414   0.0515   2e-04
##
## sigma^2 estimated as 0.0001355:  log likelihood=799.7
## AIC=-1591.4   AICc=-1591.25   BIC=-1577.1
##
## Training set error measures:
##                         ME       RMSE         MAE       MPE     MAPE
## Training set -7.879195e-05 0.01131768 0.008525257 -26.99577 189.8139
##                   MASE       ACF1
## Training set 0.5168063 0.03898039
```

Part B

How do your model compares with the one found by auto.arima(...)?

Solution

When comparing models based on AICc and BIC, it is clear that the model fit in Question 3 with the paramaters found above outperforms the auto.arima() model found in Question 2. Going foreward, this is our best model so far.

Question 4

Part A

```
# manual arima for non-seasonal
P4_arima = auto.arima(diff(BoxCox(PG1.tr,L)), d=1, D=0, max.p = 5, max.q = 3, max.P = 1, max.Q = 4)
summary(P4_arima)
```

```
## Series: diff(BoxCox(PG1.tr, L))
## ARIMA(1,1,1)(0,0,1)[12] with drift
##
## Coefficients:
##           ar1     ma1    sma1   drift
##       -0.7706  0.9100  0.6691  0.0001
## s.e.   0.0783  0.0481  0.0401  0.0018
##
## sigma^2 estimated as 0.0003157:  log likelihood=745.37
## AIC=-1480.74   AICc=-1480.53   BIC=-1462.46
##
## Training set error measures:
##                        ME       RMSE        MAE       MPE     MAPE
## Training set 1.841155e-05 0.01761193 0.01431172 -62.45225 271.8707
##                   MASE        ACF1
```

```
## Training set 1.963427 -0.01199073
```

Part B

How do your model compares with the ones found in Questions 2 and 3?

Solution

When comparing the three models so far, looking at AICc and BIC, it is clear that the model fit in Question 4, our projection with a non-seasonal difference, with the paramaters found above outperforms the auto.arima() model found in Question 2 and the seasonally differenced model found in Question 3. Going foreward, this is our best model that we have found thus far.

Question 5

```r
PG2.tr <- window(PG, end=c(2011,12))
PG2.te <- window(PG, start=c(2012,1))
```

Part A

```r
# BoxCox for PG2
L2 = BoxCox.lambda(PG2.tr)
L2
```

```
## [1] -0.3623298
```

Part B

```r
# Both differenced
both_PG2.tr = diff(diff(BoxCox(PG2.tr,L2),12))
adf.test(both_PG2.tr)
```

```
## Warning in adf.test(both_PG2.tr): p-value smaller than printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  both_PG2.tr
## Dickey-Fuller = -9.768, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

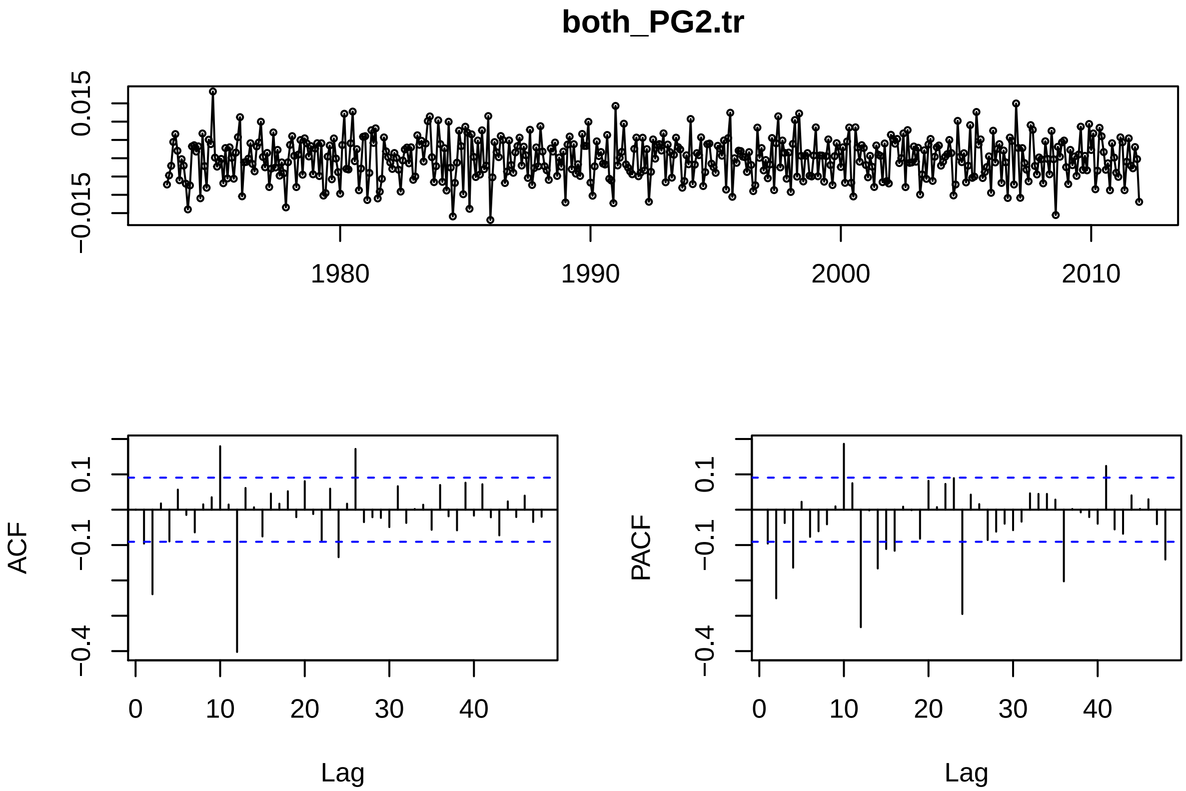
What do you conclude from the test?

Solution

What can be seen from this test is that we must reject the null hypothesis that this is a stationary time series. The p-value returned is significant and therefore allows us to say this this is not a stationary series.

Part C

```r
# Values of p, q, P, and Q
tsdisplay(both_PG2.tr,lag=48)
```




```
## p = 3, q = 3, P = 4, Q = 2
```

If you were to fit an ARIMA model to the time series you obtained above, what would be the maximum order of the (p,1,q) (P,1,Q)__12 model? (i.e., what is the maximum values of p,P,q and Q?)

Solution

Given that we are differencing both seasonally and non-seasonally (d = 1 and D = 1), the max values of the remain ARIMA parameters are: - p = 3, q = 3, P = 4, and Q = 2

Question 6

Part A

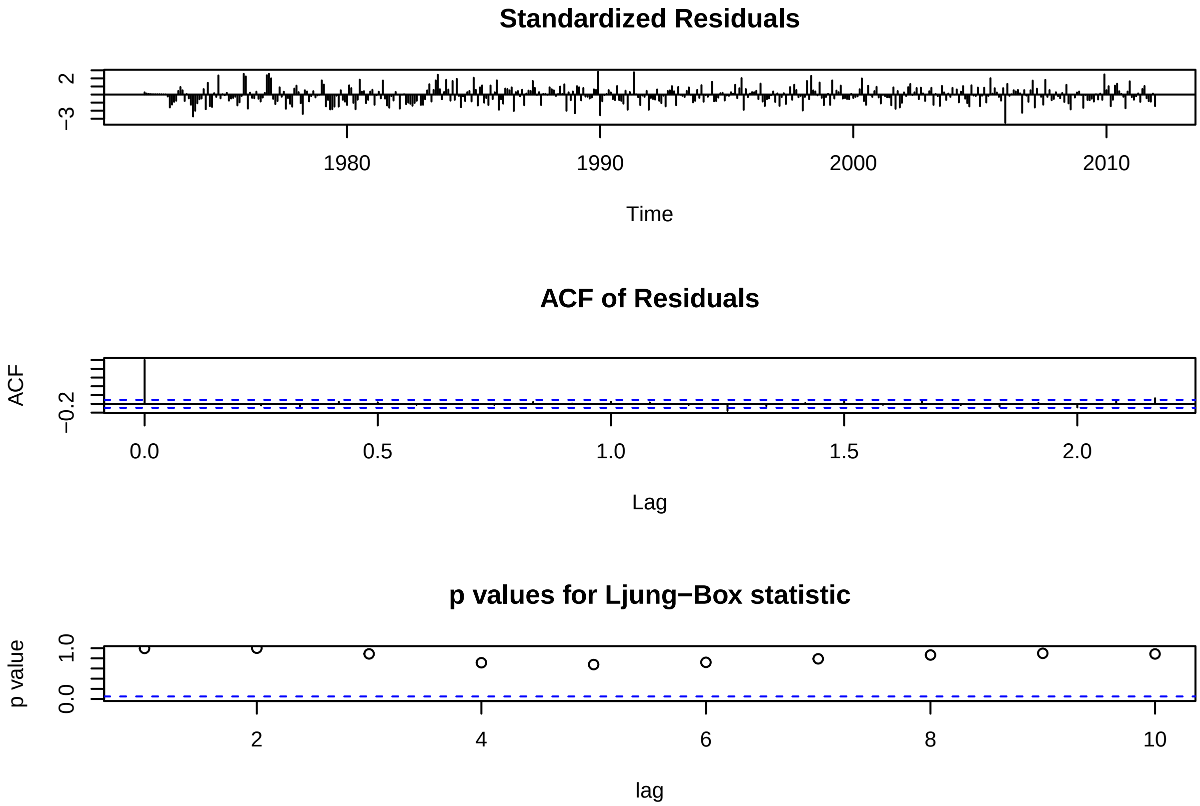
```r
# Auto arima for PG2
auto_arima2 = auto.arima(PG2.tr, lambda = L2)
summary(auto_arima2)
```

```
## Series: PG2.tr
## ARIMA(2,1,2)(0,1,1)[12]
```

```
## Box Cox transformation: lambda= -0.3623298
##
## Coefficients:
##          ar1     ar2      ma1      ma2     sma1
##       0.0494  0.2194  -0.1943  -0.5151  -0.7817
## s.e.  0.1692  0.1437   0.1556   0.1480   0.0319
##
## sigma^2 estimated as 1.878e-05:  log likelihood=1876.48
## AIC=-3740.96   AICc=-3740.78   BIC=-3716.08
##
## Training set error measures:
##                     ME     RMSE      MAE        MPE     MAPE      MASE
## Training set -0.1270695 1.559354 1.136589 -0.2102917 1.537755 0.4589151
##                   ACF1
## Training set -0.1105239
```

Part B

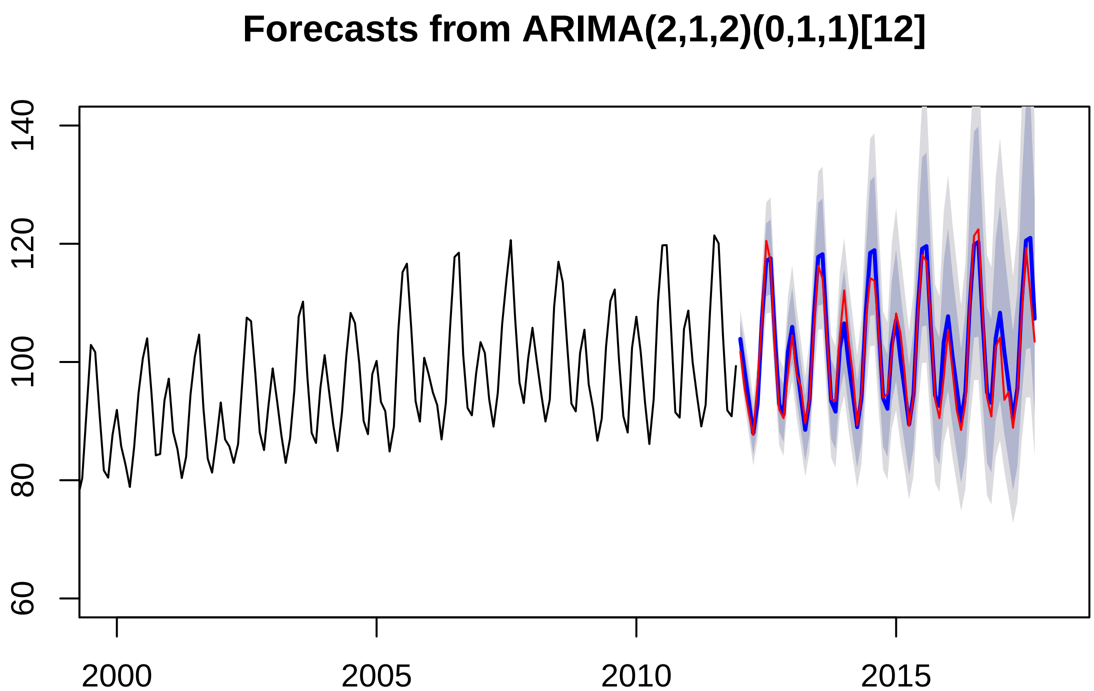
```r
# Residual of PG2
tsdiag(auto_arima2)
```



Part C

```r
# 2nd forecast
forecast_69months = forecast(auto_arima2, h = 69)
```

```r
plot(forecast_69months, xlim=c(2000,2018), ylim=c(60,140))
lines(PG2.te, col='red')
```



Part D

```r
# Check accuracy
accuracy(forecast_69months, PG2.te)
```

```
##                      ME     RMSE      MAE        MPE     MAPE      MASE
## Training set -0.1270695 1.559354 1.136589 -0.2102917 1.537755 0.4589151
## Test set     -0.8180414 2.788739 2.125495 -0.8162788 2.073699 0.8582009
##                    ACF1 Theil's U
## Training set -0.1105239        NA
## Test set      0.5016556 0.3206488
```

Based on the visual inspection of the forecast plot and the out-of-sample fit statistics comment on the forecast bias.

Solution

Similar to the previous forecast, we can see that the prediction is again below the blue testing line. With this, and seeing that the ME again grows from training to testing set, it can be see that the projection is more bias towards smaller growth than what is actual.

Question 7

Part A

```
PG3.tr <- window(PG, start=c(2005,1), end=c(2011,12))
L3 = BoxCox.lambda(PG3.tr)
L3
```

```
## [1] -0.9999242
```

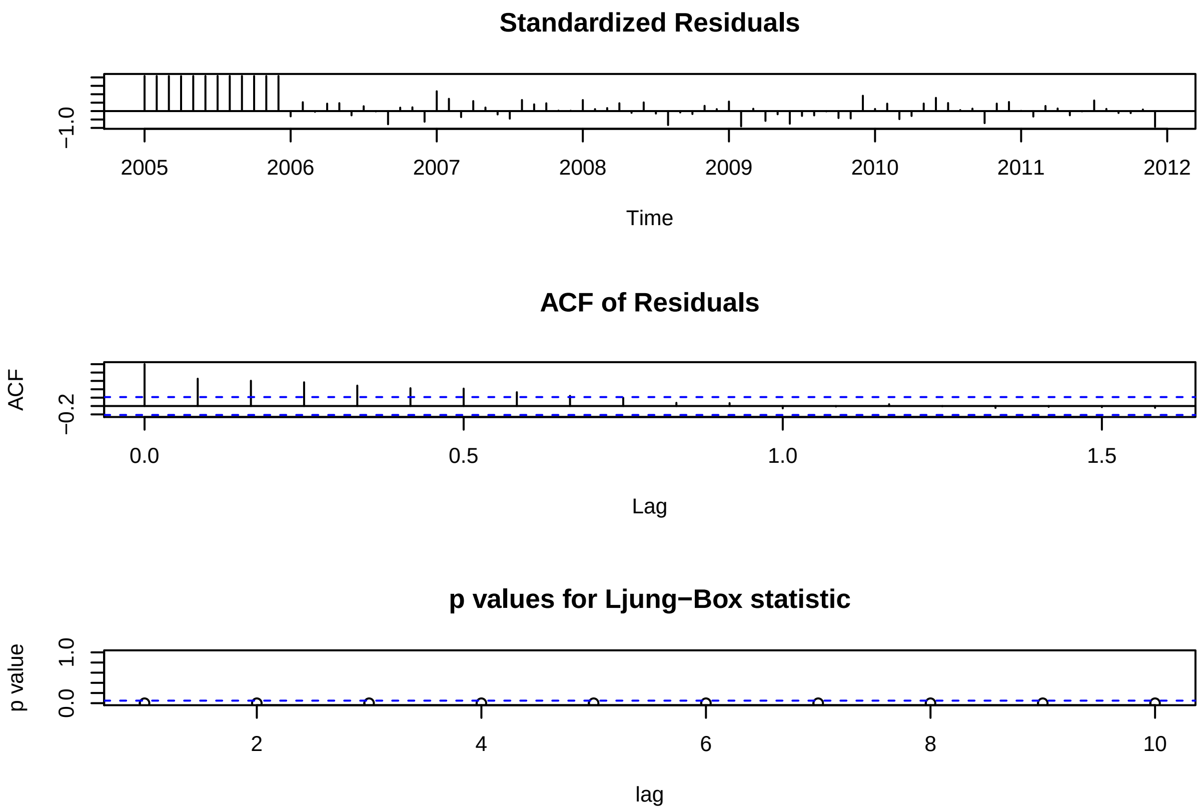
Part B

```
# Auto arima
auto_arima3 =  auto.arima(PG3.tr,lambda=L3)
summary(auto_arima3)
```

```
## Series: PG3.tr
## ARIMA(2,0,1)(2,1,0)[12]
## Box Cox transformation: lambda= -0.9999242
##
## Coefficients:
##           ar1     ar2     ma1     sar1     sar2
##       -0.2183  0.3658  0.9600  -0.5220  -0.4075
## s.e.   0.1262  0.1251  0.0716   0.1239   0.1300
##
## sigma^2 estimated as 2.262e-07:  log likelihood=501.01
## AIC=-990.01   AICc=-988.72   BIC=-976.35
##
## Training set error measures:
##                    ME     RMSE      MAE      MPE     MAPE     MASE
## Training set 1.496384 4.015099 2.812543 1.469976 2.790363 1.010773
##                   ACF1
## Training set 0.6004094
```

Part C

```
# Residuals of arima 3
tsdiag(auto_arima3)
```



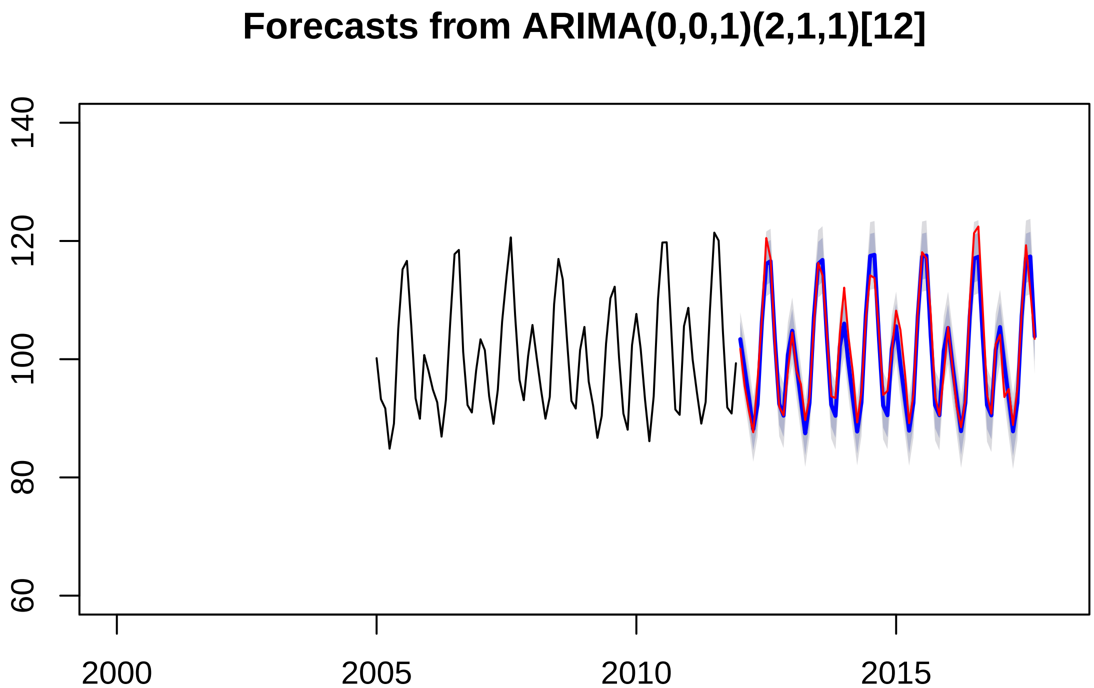
Part D

```r
# Create best model
best_arima = auto.arima(PG3.tr, max.d=2, max.D=2, max.p = 1, max.q = 1, max.P = 2, max.Q = 1)
summary(best_arima)
```

```
## Series: PG3.tr
## ARIMA(0,0,1)(2,1,1)[12]
##
## Coefficients:
##          ma1     sar1    sar2     sma1
##       0.6714  -0.0817  -0.252  -0.6307
## s.e.  0.0867   0.2992   0.197   0.3639
##
## sigma^2 estimated as 5.376:  log likelihood=-166.48
## AIC=342.96   AICc=343.87   BIC=354.34
##
## Training set error measures:
##                     ME     RMSE     MAE       MPE     MAPE      MASE
## Training set 0.4052349 2.086235 1.57698 0.3751014 1.551031 0.5667358
##                    ACF1
## Training set 0.06007128
```

Part E

```r
# Best fit model forecast
best_forecast = forecast(best_arima, h = 69)
plot(best_forecast, xlim=c(2000,2018), ylim=c(60,140))
lines(PG2.te, col='red')
```




```r
accuracy(best_forecast, PG2.te)
```

```
##                     ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 0.4052349 2.086235 1.576980 0.3751014 1.551031 0.5667358
## Test set     0.7863188 2.649373 2.067733 0.7602532 2.013650 0.7431029
##                    ACF1 Theil's U
## Training set 0.06007128        NA
## Test set     0.41441393 0.3048199
```

Part F

Based on the visual inspection of the forecast plot and the out-of-sample fit statistics comment on the forecast bias.

Solution

What can be seen here is that the best forecast is fitting the testing data very similarly to the training set. The RMSE is around the same for both of these, plus the bias is not as large as in previous models. This forecast is a slight over-approxiamation of the data as can be seen by the prediction line that is slightly higher than the testing data.

Part G

Compare the best model you obtained for the PG3.tr training set (this question) with the model you obtained for PG2.tr training set (Question 6) and comment on their out-of-sample fit statistics. Explain why you cannot compare the AICc and BIC of both models?

Solution

This model seems to have the best fit in both in and out of sample statistics. It produced the lowest AICc and BIC for our data, and therefore will be used in the next question as the best-fitting prediction model to forecast out.

Question 8

Part A

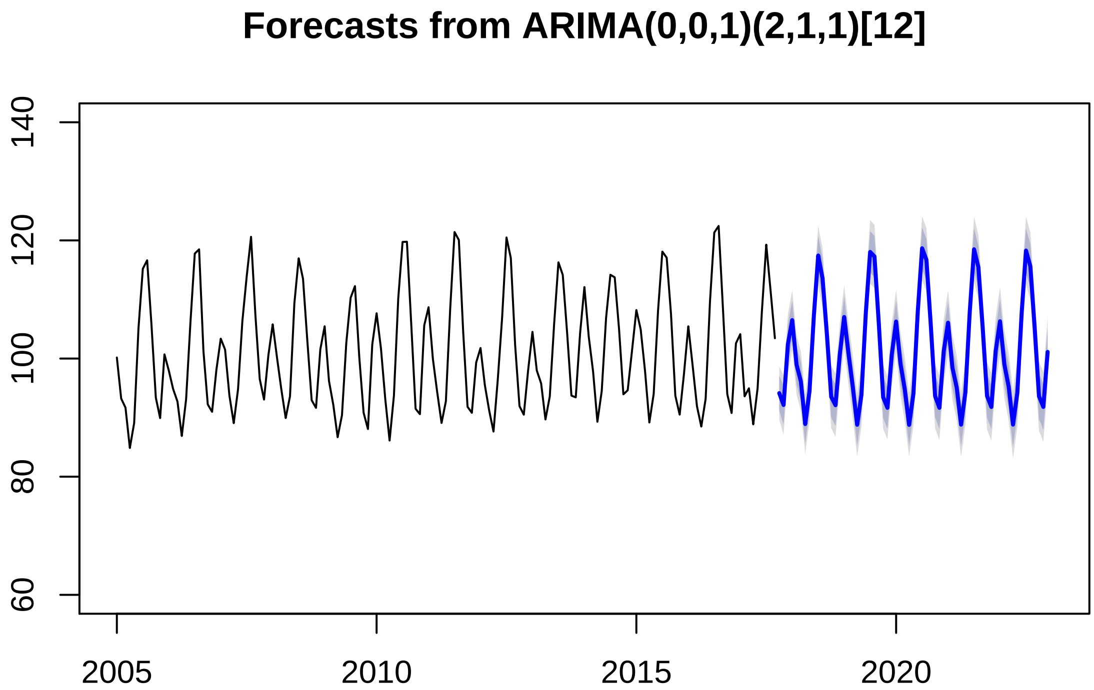
```
PG.tr <- window(PG, start=c(2005,1))
```

Part B

```
# Arima to project the next 63 months
best_arima_fit = arima(PG.tr, order = c(0,0,1), seasonal=c(2,1,1))
summary(best_arima_fit)
```

```
##
## Call:
## arima(x = PG.tr, order = c(0, 0, 1), seasonal = c(2, 1, 1))
##
## Coefficients:
##          ma1    sar1     sar2     sma1
##       0.4967  0.0369  -0.3275  -0.6647
## s.e.  0.0637  0.1164   0.1020   0.1073
##
## sigma^2 estimated as 5.276:  log likelihood = -323.7,  aic = 657.39
##
## Training set error measures:
##                     ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 0.3986409 2.205279 1.710522 0.3679101 1.666352 0.2302728
##                    ACF1
## Training set 0.07384247
```

```
final_forecast = forecast(best_arima_fit, h = 63)
plot(final_forecast, xlim=c(2005,2023), ylim=c(60,140))
```


Forecasts from ARIMA(0,0,1)(2,1,1)[12]
60
80
100
120
140
2005
2010
2015
2020