**Assignment #9**

**Seasonal ARIMA Laboratory**

Due: Wed. Nov. 8 (by 11:59pm)

(40 pts. Total)

In this assignment we will focus on estimation of electricity generation in the US through December 2022 (63 months into the future). We will base our analysis on the monthly data (index) provided by the Federal Reserve in

I have downloaded this data in the CSV file named “IPG2211N.csv”. The following sequence of instructions set up the data in your R-Studio Global Environment, and creates the time-series variable PG so you can start your assignment.

**library**(fpp2)

**library**(dplyr)

PG <- **read.csv**("IPG2211N.csv") %>%  
 **select**(-DATE) %>%  
 **ts**(start=**c**(1972,1), frequency=12)

Initially we will set up as training data the series from January 1972 through December 1995, and the testing set as the data from January 1996 through December 2000. This is accomplished with the following two commands:

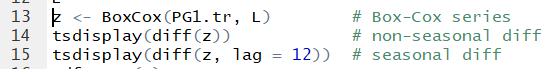
PG1.tr <- **window**(PG, end=**c**(1995,12))  
PG1.te <- **window**(PG, start=**c**(1996,1), end=**c**(2000,12))

1. (5 pts.) Preliminary analysis of training data:
   * Obtain the Box-Cox transformation parameter lambda for the training set PG1.tr

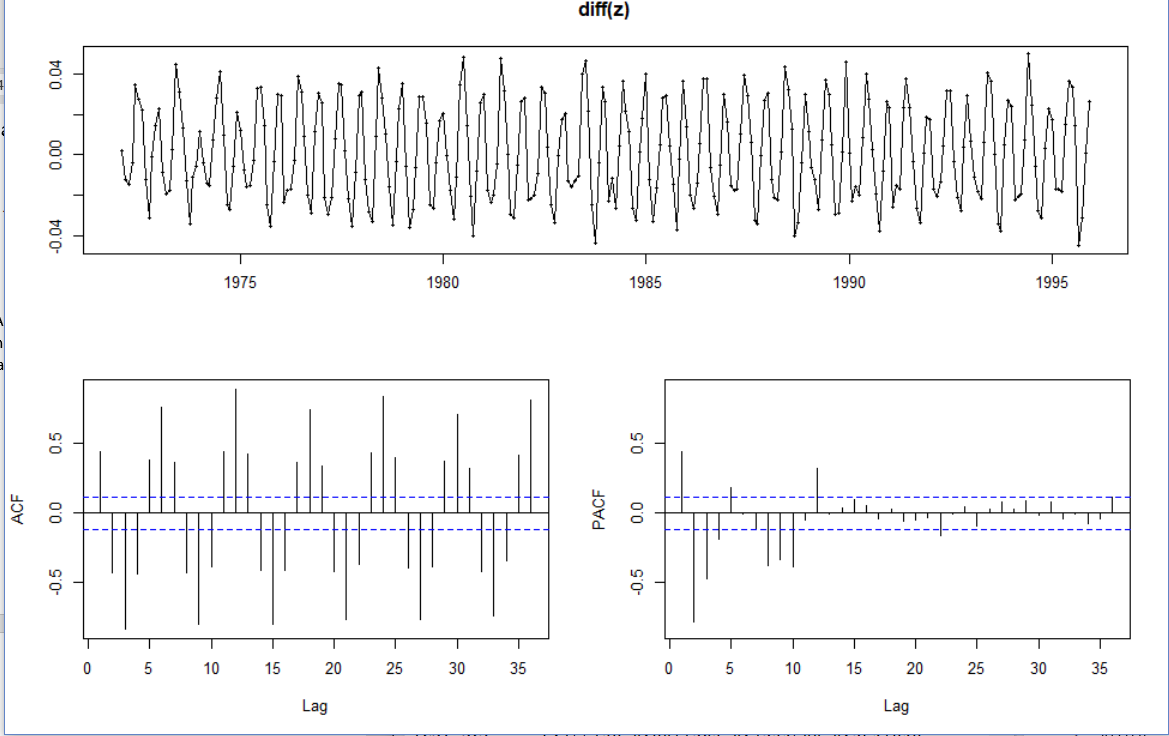


L is -0.254.

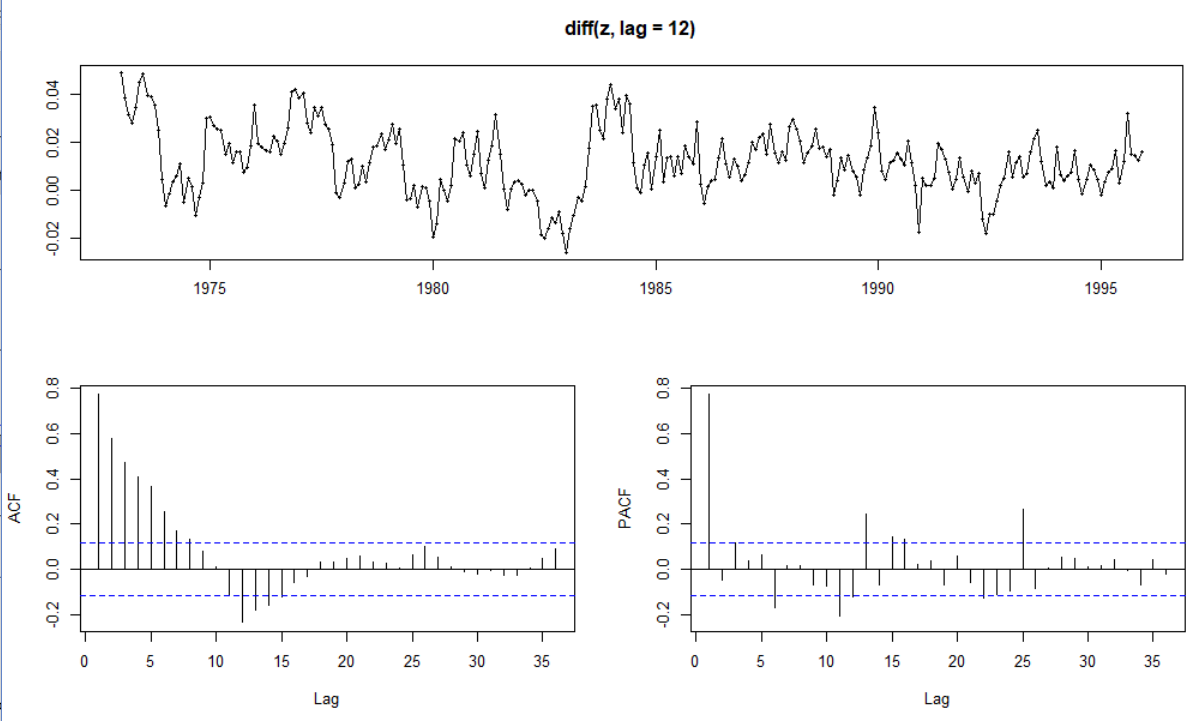
* + Use the tsdisplay(…) function to examine ACF and PACF of the Box-Cox transformed series with a seasonal difference and a non-seasonal difference. Do the differenced and transformed series look stationary?



Non-seasonal difference ACF and PACF:

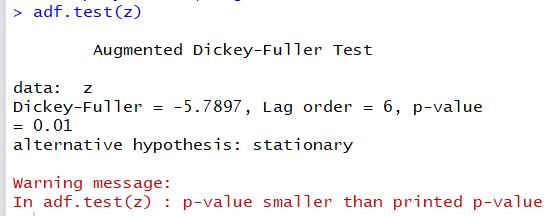


Seasonal difference ACF and PACF:



The non-seasonal differences do not look seasonal. It is not white noise. The differenced and transformed series look stationary from the plots of the data. There is evidence of seasonality from looking at the differencing at the non-seasonal level. The PACF drops to a non-significant value fairly quickly, while the ACF values are decreasing exponentially with the seasonal difference and showing a seasonal pattern with the non-seasonal difference. The series look fairly stationary, but the ACF and PACF plots are not perfectly clean.

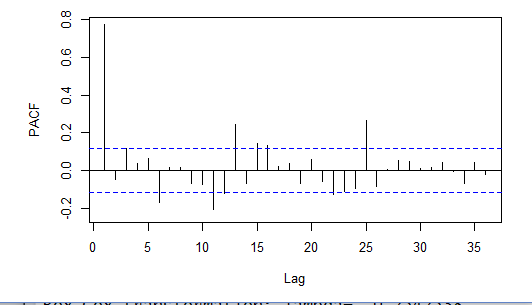
* + Run the adf.test(…) on the above series. What do you conclude from the test?



The adf.test function showed a p-value of less than 0.01, meaning that the series passed the test that it is stationary. We reject the null hypothesis that it is non-stationary and instead accept the alternative hypothesis that it is a stationary time series.

* + If you were to fit an autoregressive model to the non-seasonal component of the time series, what would be the maximum order of this model? (i.e., what is the maximum value of that you would consider for ?

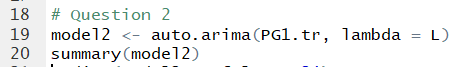
AR(x) models are fitted when the PACF cuts off sharply and the ACF shows some exponential decay. From examining the tsdisplay(diff(z, lag = 12)) PACF graph, we see that the PACF cuts off sharply after lag = 1. Thus, the maximum value of p that I would consider for an AR() model is 1.



* + What would be the maximum number of that you would consider for model ?

In order to determine the maximum value of P, I will look for significant spikes at lag 12, 24, 36, … in the PACF and exponential decay in the seasonal lags of the ACF. The PACF shows a barely significant spike at lag = 12, and none of the other multiples of 12 lags are significant. In the ACF graph, the seasonal lags at 12 are decaying exponentially. Because the significant spike is only at lag = 12, the maximum number of P I would consider for this model is 1.

1. (5 pts.) Automatic ARIMA model selection:
   * Run the auto.arima(…) function to fit an ARIMA model on the Box-Cox transformation of the PG1.tr dataset, and report the order of the model, the value of the model parameters and the value of the AICc and BIC information criteria?



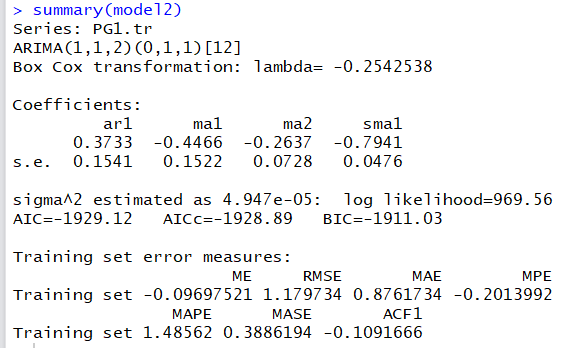
The order of the model is (1,1,2)(0,1,1)12

Model parameters:

* ar1: 0.3733
* ma1: -0.4466
* ma2: -0.2637
* smal: -0.7941
* sigma2: .00004947

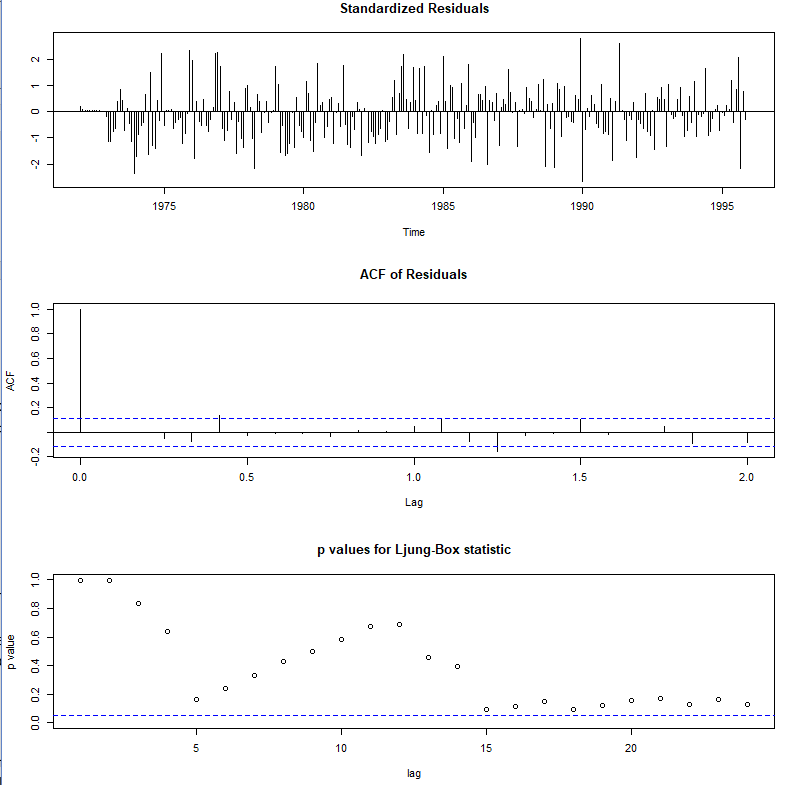
AICc: -1929.12

BIC: -1911.03



* + Use the function tsdiag(…, gof.lag=24) to assess the validity of the model you obtained in Question 1. Based on the results you obtained comment on the validity of the model.

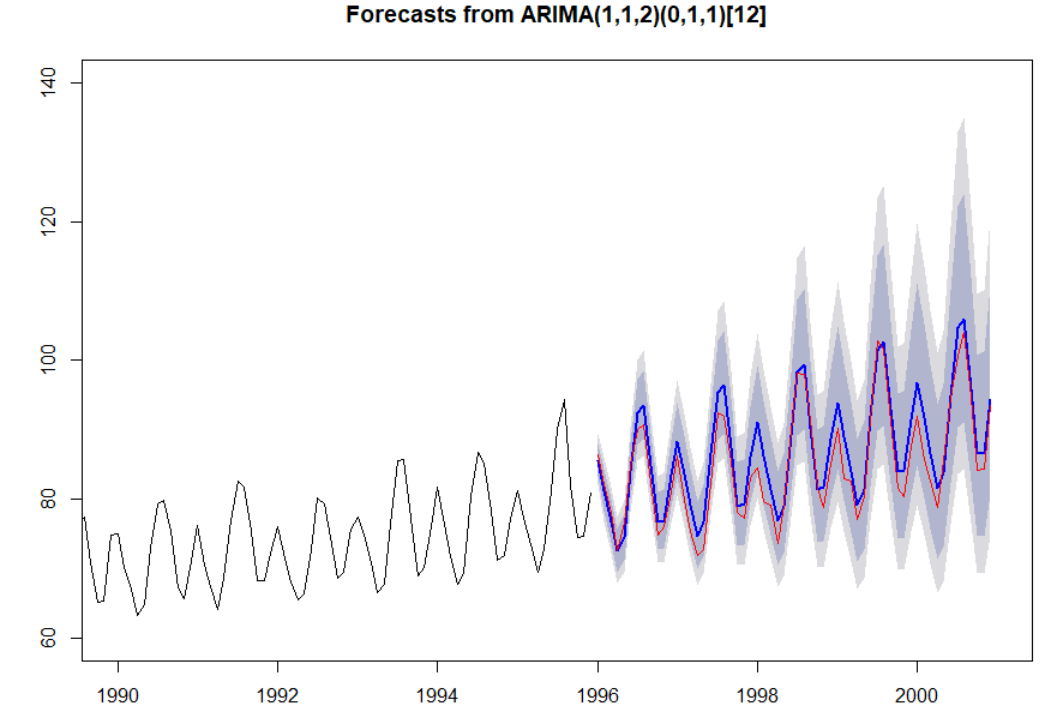




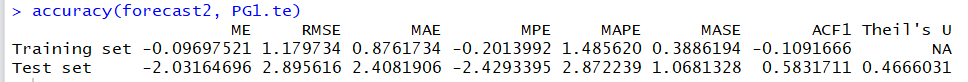
According to the standardized residuals, this model is valid—all of the values are between -3 and +3. One of the ACF values is significant at lag = 0.4, meaning the model is barely not valid. The p-values all pass the validity test, as all of them are non-significant. However, we see issues with a pattern in the p-values and the p-values after lag = 15 being very close to significance. The points after lag = 15 may not be stationary. While this model may be “valid,” the ACF and p-values show that it is not the best choice for correctly fitting the data. This model is not very clean.

* + Use the forecast(…) function to prepare a 60 month-ahead forecast for the electricity generation and then overlay (using a red line) the actual data for electricity generation. To examine visually the forecast in greater detail use xlim=c(1990, 2001), and ylim=c(60,140) in your forecast plot.





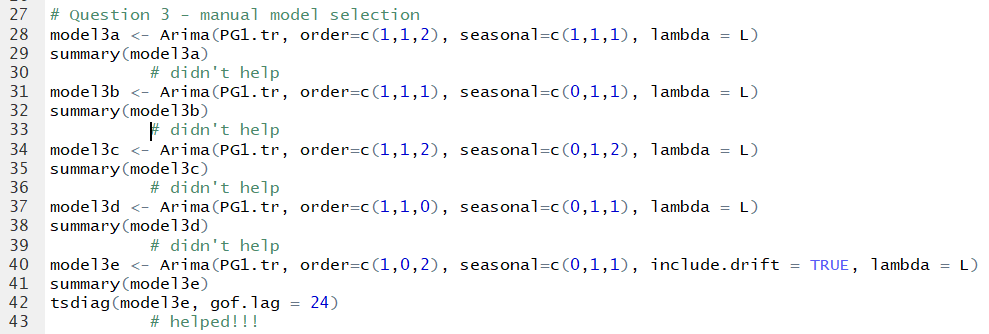
* + Use the accuracy(…) function to obtain the training and testing fit (PG1.te) metrics for the model obtained. Based on the visual inspection of the forecast plot and the out-of-sample fit statistics comment on the forecast bias.

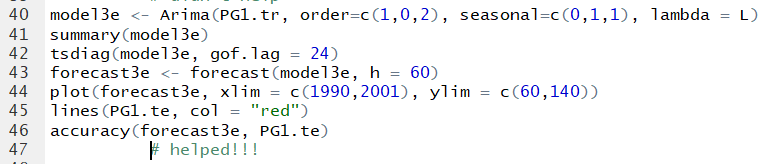


Based on the visual inspection of the forecast plot and the out-of-sample fit statistics, this is not a very good model for the data. The forecast (the blue line) is consistently higher than the actual values shown in red. The confidence intervals for these predictions are also very wide. From visual inspection, this model does not forecast the values very well. The out-of-sample MASE is 1.068, which is not very good. That means that there is about 107% mean absolute scaled error. The MAPE is much worse, at 2.87 mean absolute percentage error. With these out-of-sample fit statistics, we see than the forecast is quite biased.

1. (5 pts.) Manual Model Selection:
   * Search manually for a model to improve on the automatic selection in Question 2. To limit your manual search do not exceed the maximum values of and that you identified in Question 1.

The maximum values of p and P that I identified in Question 1 were 1. I tested possible values of d, q, D, and Q to find a model to improve on the automatic selection from Question 2.



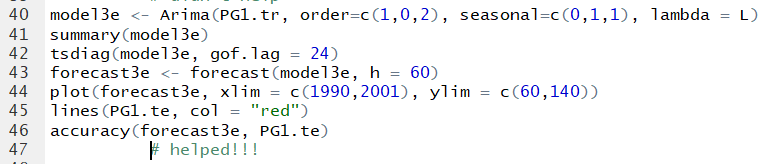


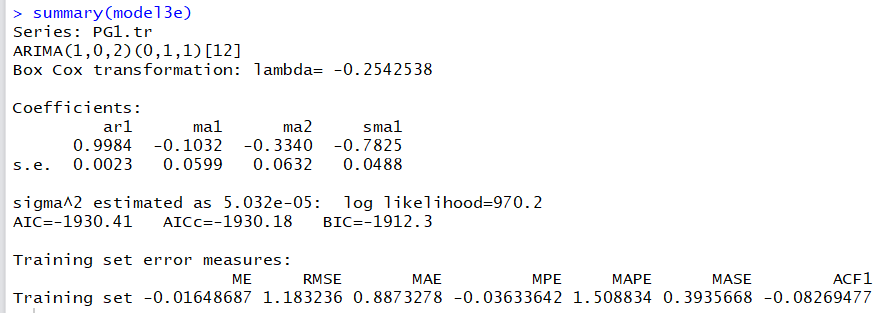
* + Report on the best model that you identified and comment on its AICc and BICc. Does your model have any out-of-sample bias problem?

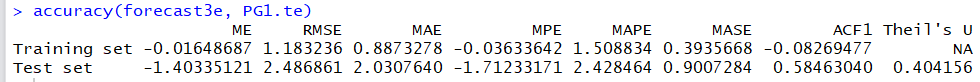
The auto.arima selection from Question 2 was a (1,1,2)(0,1,1)12 model with an AICc of -1928.89, BIC of -1911.03, MAPE of 2.87, and a MASE of 1.07.

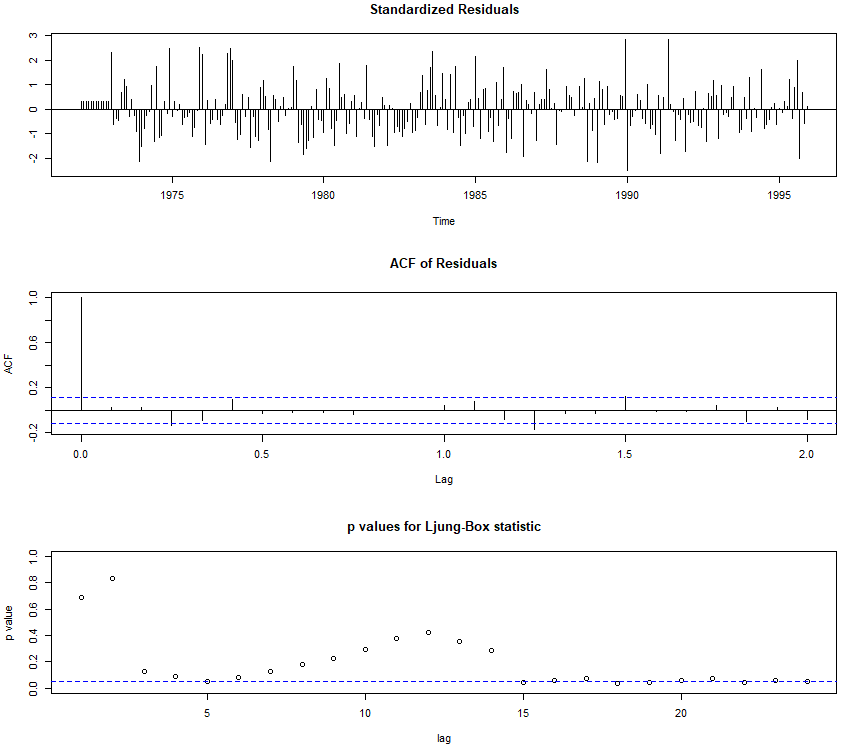
The best model that I identified was an ARIMA (1,0,2)(0,1,1)12 model with an AICc of -1930.18 and a BIC of -1912.3. These AICc and BIC values were slightly lower than the ones from the auto.arima model.

There are some issues with the bias and stationarity of this model. The tsdiag plots show that the p-values are nearly significant and there are some non-significant ACF values, meaning that this model is not perfectly stationary. The out-of-sample fit statistics are also still high, even though they improve on the automatic selection model. The MAPE was reduced from 2.87 to2.42, and the MASE was reduced from 1.07 to 0.9. These numbers, though smaller, still show a significant amount of out-of-sample bias in the forecast model.





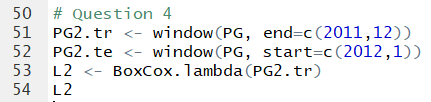




1. (5 pts.) ARIMA model for the expanded training set:
   * Now we redefine the training and testing sets as follows:

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

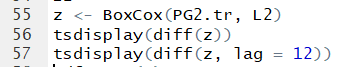
* + Obtain the Box-Cox transformation parameter lambda for the training set PG2.tr



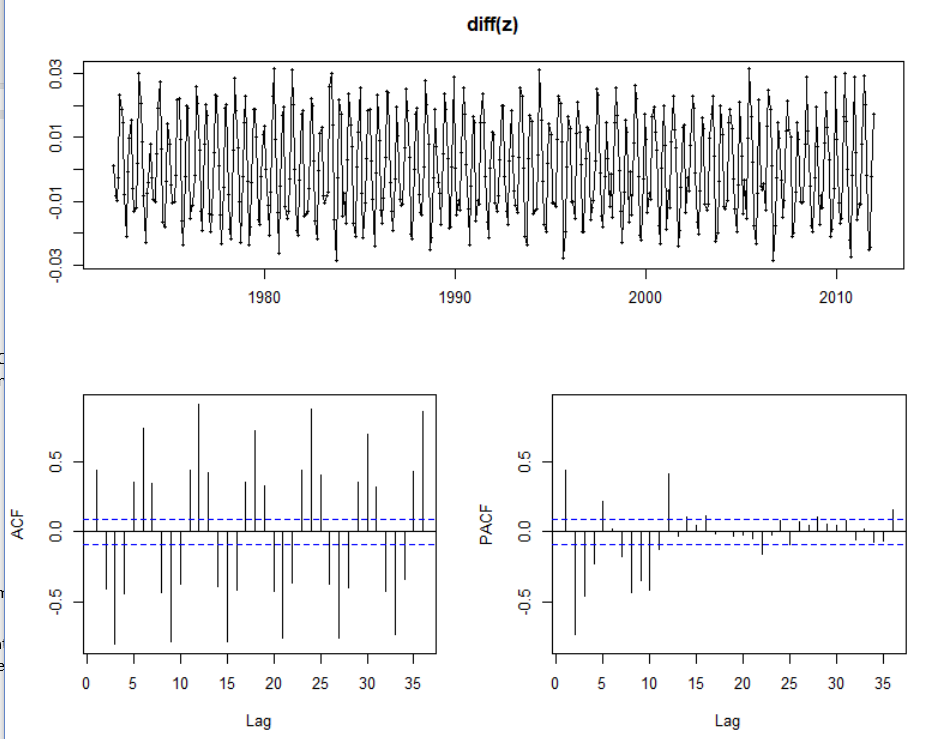


Lambda for the expanded training set is -0.362.

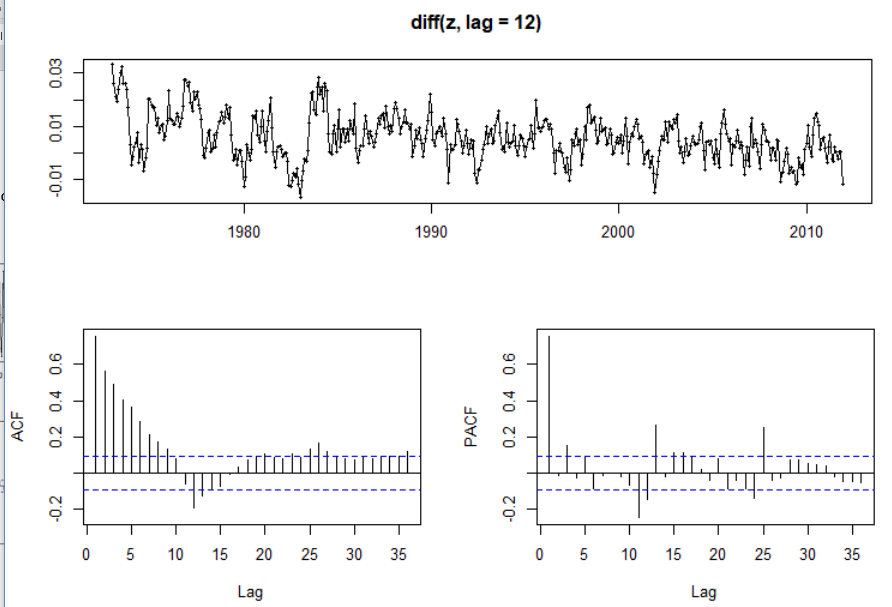
* + Use the tsdisplay(…) function to examine ACF and PACF of the Box-Cox transformed series with a seasonal difference and a non-seasonal difference. Do the differenced and transformed series look stationary?



Tsdisplay of Box-Cox transformed series with a non-seasonal difference:

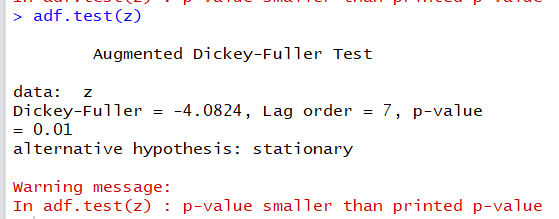


Tsdisplay of Box-Cox transformed series with a seasonal difference:



The differenced and transformed series look less stationary than in question 1. The PACF values take a while to become non-significant. However, the plots of the data do look stationary. Both the non-seasonal difference and the seasonal difference plots have values past lag 10 that are significant.

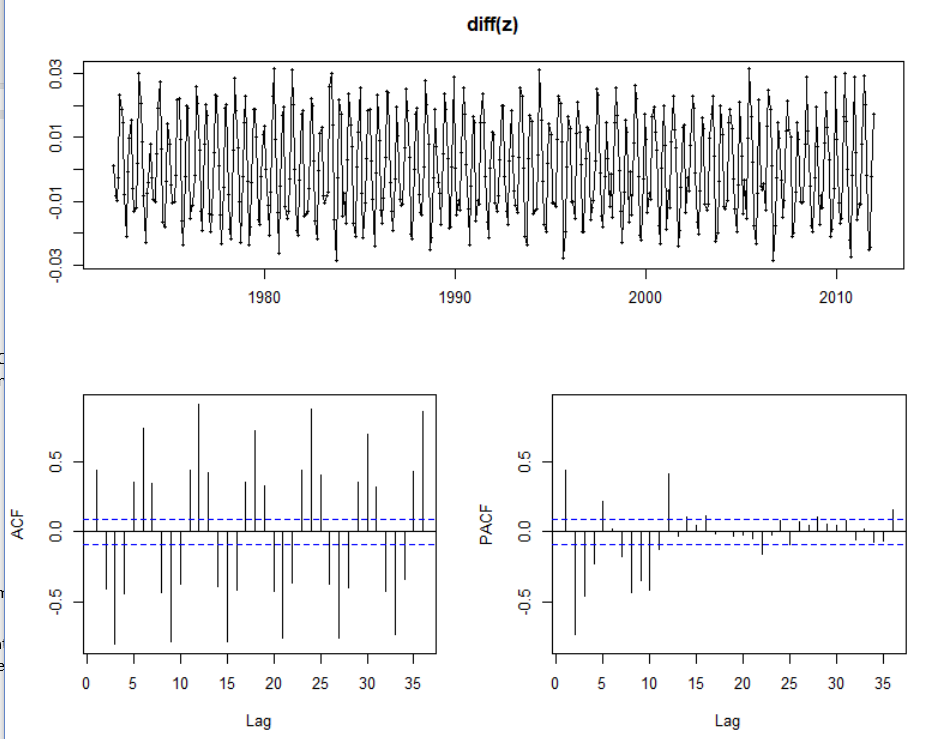
* + Run the adf.test(…) on the above series. What do you conclude from the test?



The adf test shows a p-value of less than 0.01, meaning that we can reject the null hypotheses that the data is non-stationary and accept the alternate hypothesis that the data is stationary. We conclude that this data is acceptable to run an ARIMA model on, as it passes the stationarity test.

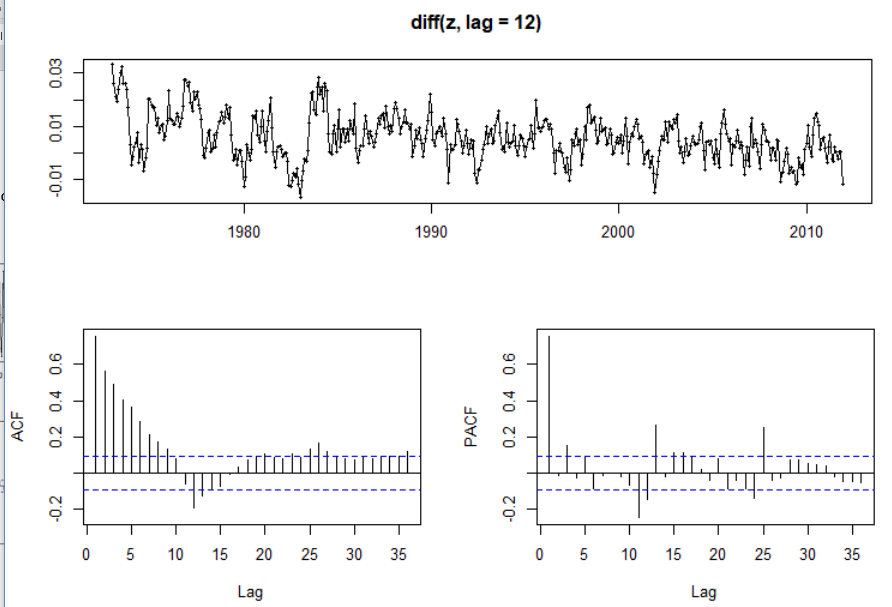
* + If you were to fit a moving average model to the non-seasonal component of the time series, what would be the maximum order of this model? (i.e., what is the maximum value of that you would consider for ?

MA(x) models are selected when the ACF drops off sharply and the PACF shows exponential decay. In the diff(z) ACF and PACF, it is not clear where the ACF drops off. After lag = 3, there is a spike in decay, so I would think that maximum order of this model would be MA(3). The maximum value of q I would pick for the model is 3.



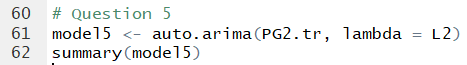
* + What would be the maximum number of that you would consider for model ?

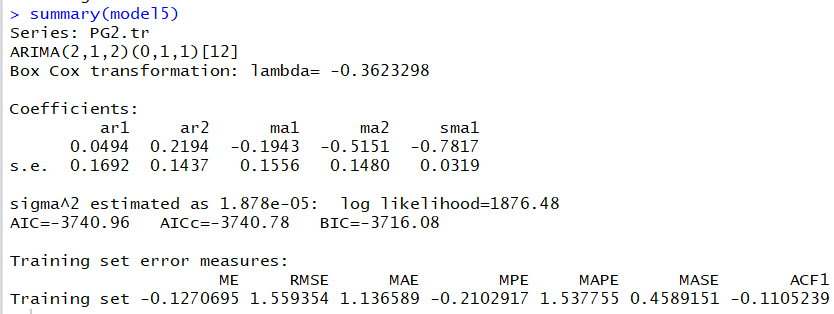
MA(x) models for seasonal components are fitted by looking for significant spikes at lag = 12, 24, 36, etc on the ACF graph and an exponential decay in the seasonal spikes on the PACF graph.



The ACF graph shows significant spikes at lag = 12, 24, and 36. The spikes become non-significant on the PACF graph after lag = 24. There are 3 significant spikes on the PACF graph at 0, 12, and 24, so the maximum value of Q that I would consider for this model would be 3.

1. (5 pts.) Automatic ARIMA model selection on the expanded dataset:
   * Run the auto.arima(…) function to fit an ARIMA model on the Box-Cox transformation of the PG2.tr dataset, and report the order of the model, the value of the model parameters and the value of the AICc and BIC information criteria?





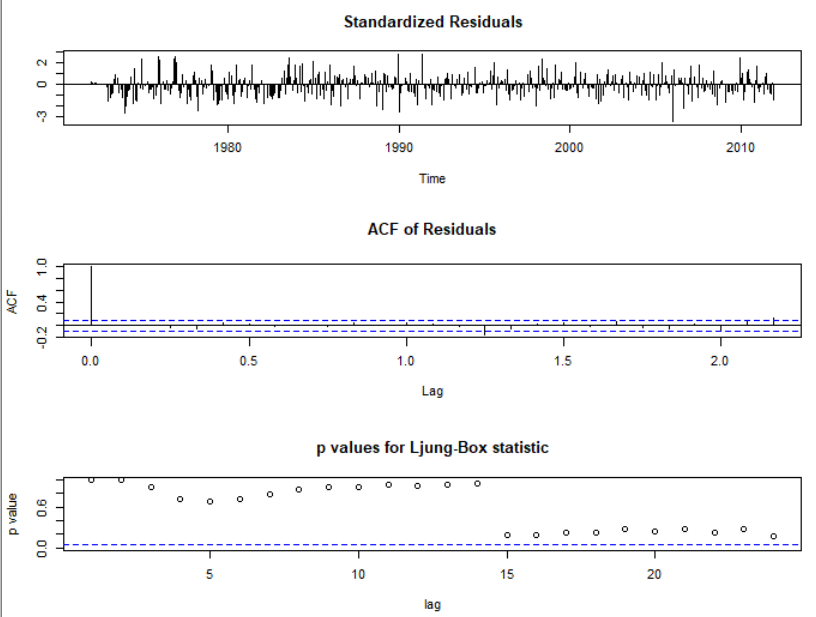
The order of the model is (2,1,2)(0,1,1)12

The values of the model parameters are

* ar1: 0.0494
* ar2: 0.2194
* ma1: -0.1943
* ma2: -0.5151
* smal: -0.7817
* sigma2: .00001878

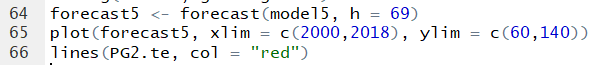
The value of the AICc is -3740.96 and the value of the BIC is -3716.08.

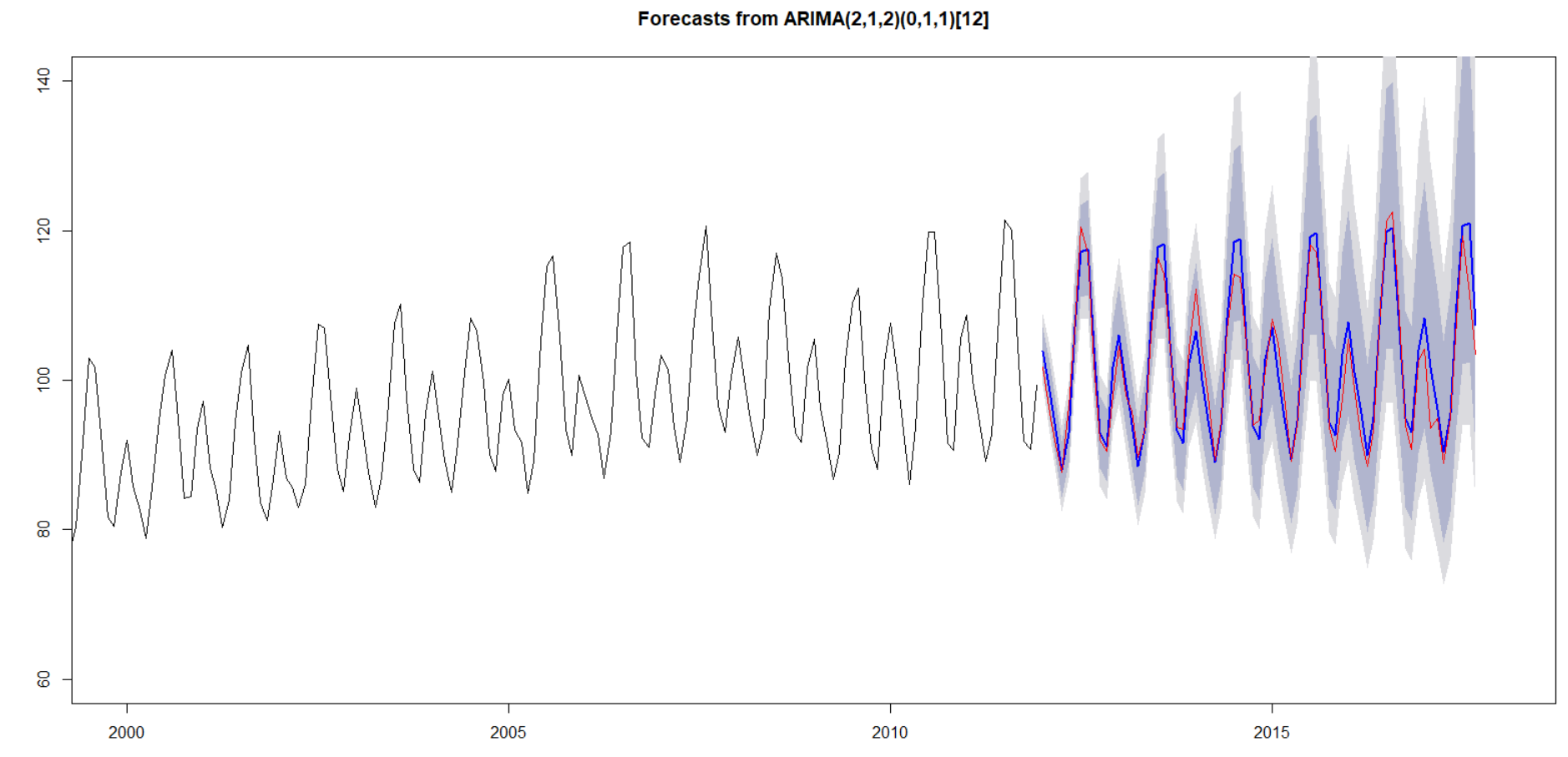
* + Use the function tsdiag(…, gof.lag=24) to assess the validity of the model you obtained. Based on the results you obtained comment on the validity of the model.



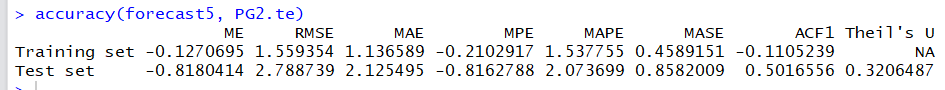
Above are the results of tsdiag(model5, gof.lag = 24). The model is valid from the standardized residuals as they are all between -3 and +3. The ACF of residuals is nearly clean. There are 2 non-significant values towards the later lags. All of the p-values are non-significant, but there is an issue after lag = 15. The p-values are very close to being significant. Though the model is technically valid, it is not very clean and does not fit the later training data well.

* + Use the forecast(…) function to prepare a **69 month-ahead** forecast for the electricity generation and then overlay (using a red line) the actual data for electricity generation. To examine visually the forecast in greater detail use xlim=c(2000, 2018), and ylim=c(60,140) in your forecast plot.





* + Use the accuracy(…) function to obtain the training and testing fit (PG2.te) metrics for the model obtained. Based on the visual inspection of the forecast plot and the out-of-sample fit statistics comment on the forecast bias.

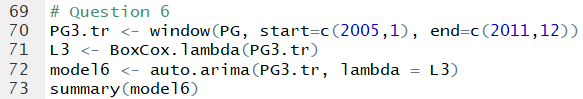


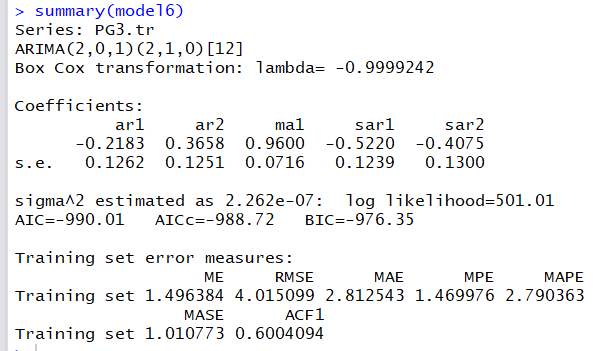
Based on the visual inspection of the forecast plot, the forecast is still very biased. The forecast is neither consistently higher or lower than the actual values, but the confidence intervals are very wide. Based on the out-of-sample fit statistics, the forecast is also very biased. The MAPE is 2.07, which shows a 207% mean absolute percentage error. The MASE is slightly better, but it still shows an 85.8% mean absolute scaled error. This forecast does not fit the data that well.

1. (5 pts.) Automatic ARIMA model selection with a reduced training dataset:
   * As the patterns of consumption and generation changed substantially on 2005, before setting on a forecasting model we will try reducing the training set to information posterior to 2005. To this end we define the training data set as follows:

PG3.tr <- **window**(PG, start=**c**(2005,1), end=**c**(2011,12))

* + Now run the auto.arima(…) function to fit an ARIMA model on the Box-Cox transformation of the PG3.tr dataset, and report the order of the model, the value of the model parameters and the value of the AICc and BIC information criteria?





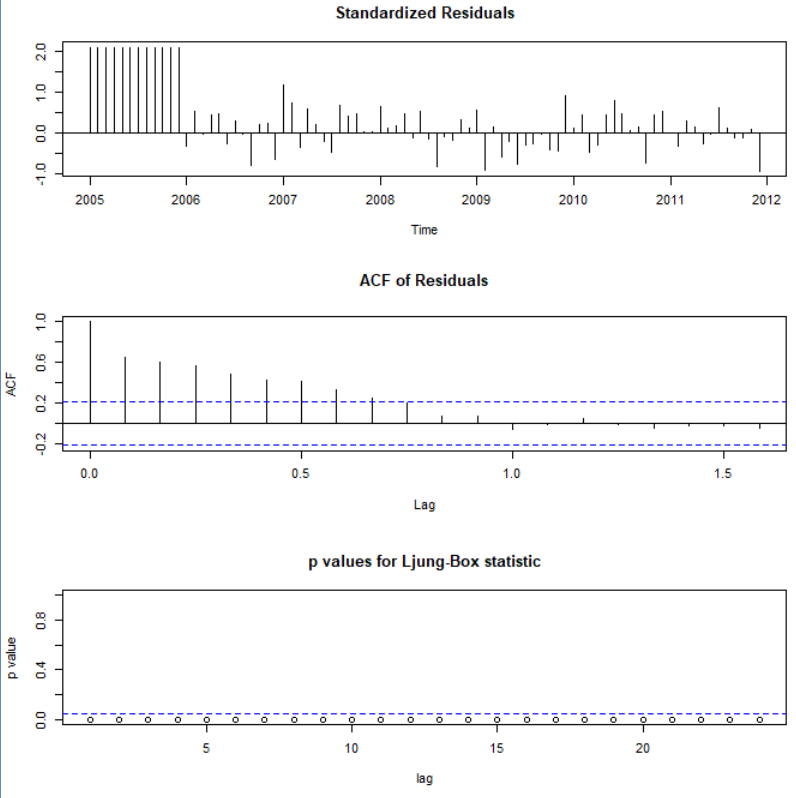
The order of the model is (2,0,1)(2,1,0)12

The model parameters are:

* ar1: -0.2183
* ar2: 0.3658
* ma1: 0.9600
* sar1: -0.5220
* sar2: -.4075
* sigma2: .000000226

The AICc is -988.72 and the BIC is -976.35.

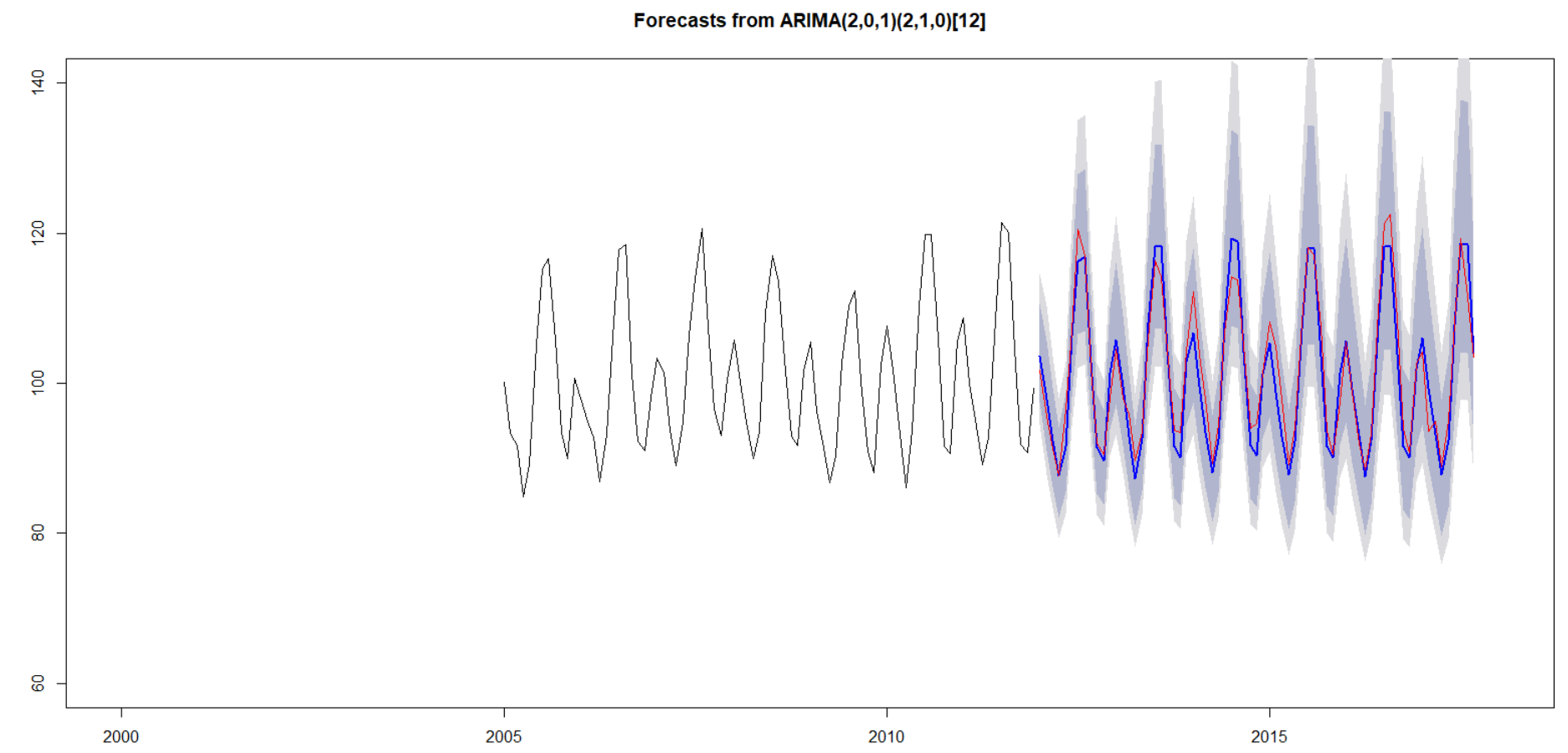
* + Use the function tsdiag(…, gof.lag=24) to assess the validity of the model you obtained in Question 1. Based on the results you obtained comment on the validity of the model.



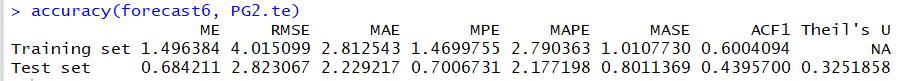
The model is valid based on the standardized residuals, as they are all between -3 and +3. However, the ACF plot and the p-value plot shows that this model is not valid. The ACF values of the residuals have more than lag = 0 as non-significant, where they should all be significant after lag = 0. The p-values should all be non-significant, and they are all significant. This model does not look valid.

* + Use the forecast(…) function to prepare a **69 month-ahead** forecast for the electricity generation and then overlay (using a red line) the actual data for electricity generation. To examine visually the forecast in greater detail use xlim=c(2000, 2018), and ylim=c(60,140) in your forecast plot.



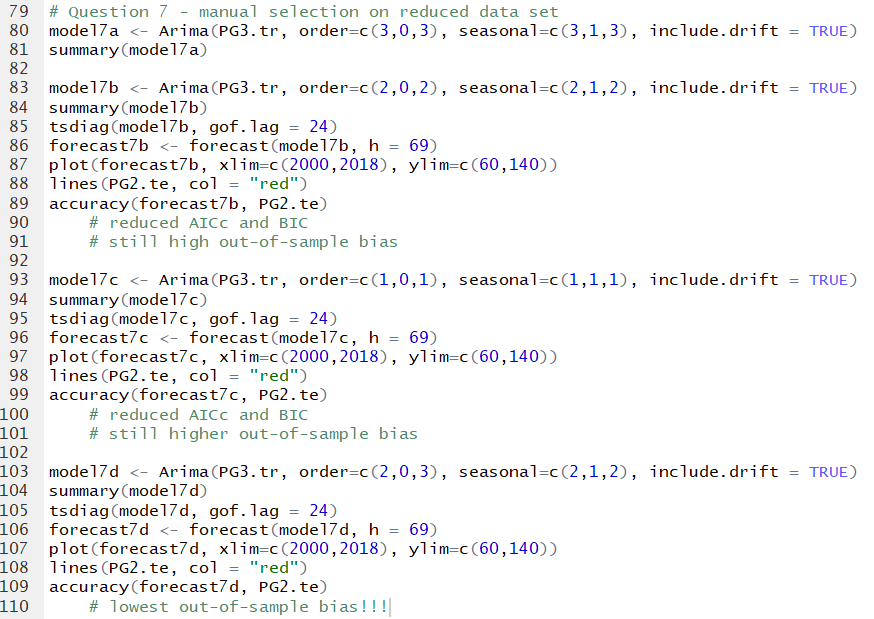


* + Use the accuracy(…) function to obtain the training and testing fit (PG2.te) metrics for the model obtained. Based on the visual inspection of the forecast plot and the out-of-sample fit statistics comment on the forecast bias.



Based on the visual inspection of the forecast plot, this model does not fit the data that well. While the forecast is not typically higher than or lower than the predicted values, the confidence intervals are quite wide. The out-of-sample fit statistics confirm this bias. The Mean Absolute Percentage error is 217%, and the Mean Absolute Scaled Error is 80%. This model shows high forecast bias, though it is the least level of forecast bias we have seen in all models so far.

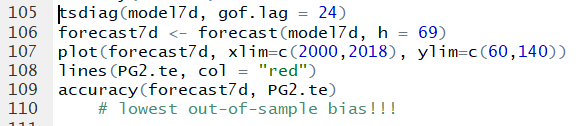
1. (5 pts.) Manual ARIMA model selection with a reduced training dataset:
   * Now use trial and error to find the best ARIMA to PG3.tr with values for ,and smaller than or equal to 3. What is the best model you could find manually?

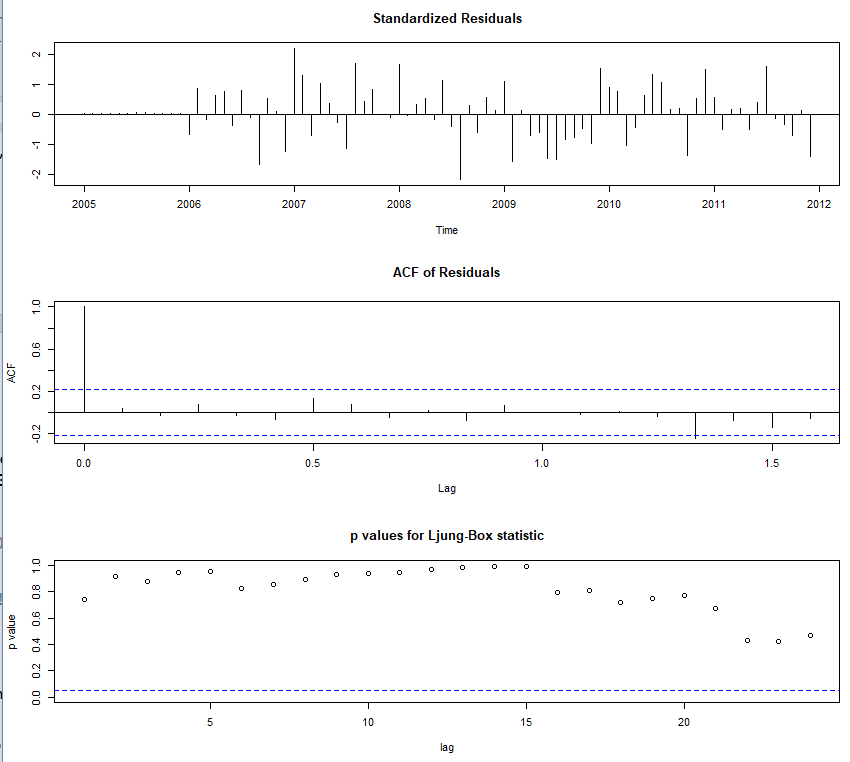


I decided to focus on the AICc and BIC values at first. However, I changed my mind to focus on the lowest out-of-sample bias (MAPE and MASE values). The code just shows some of my trials; I tried many different combinations of p, P, q, and Q in order to have the best out-of-sample bias.

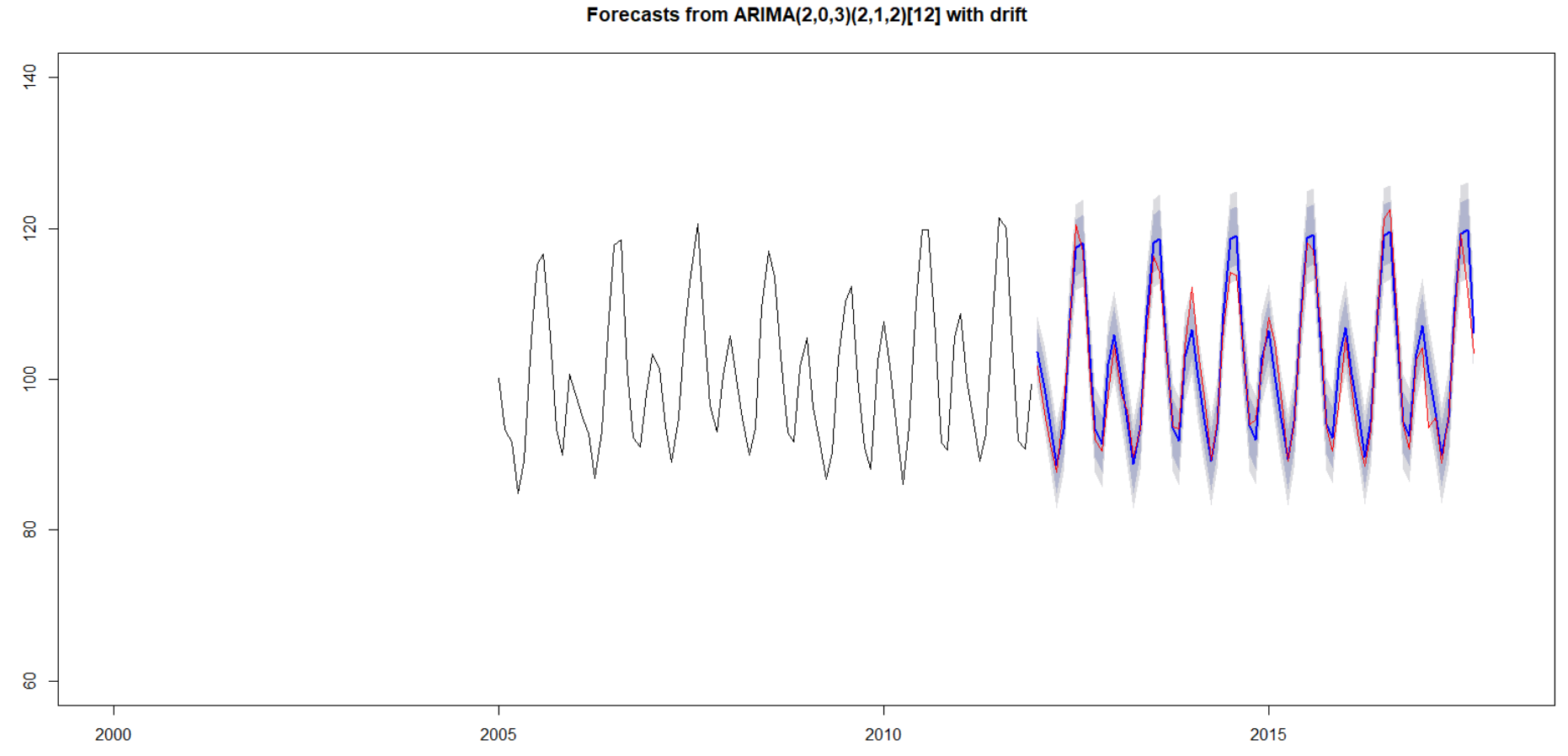
The best model I could was an ARIMA (2,0,3)(2,1,2) model with drift and no lambda transformation added.

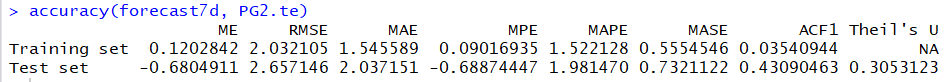
* + Run the model diagnostics and comment on the out-of-sample fit and bias when you use the model to predict the testing dataset PG2.te





The model diagnostics show that this model is valid, with the exception of one point on the ACF plot. All the standardized residuals are between -3 and +3. All of the ACF values of residuals are significant with the exception of one near lag = 1.3. All of the p-values are non-significant and not close to 0.

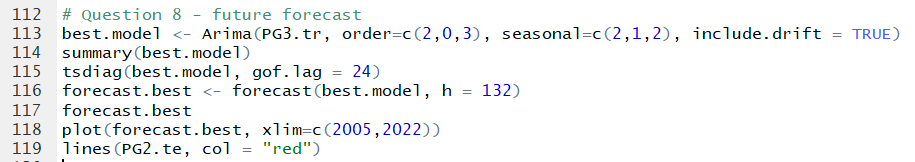




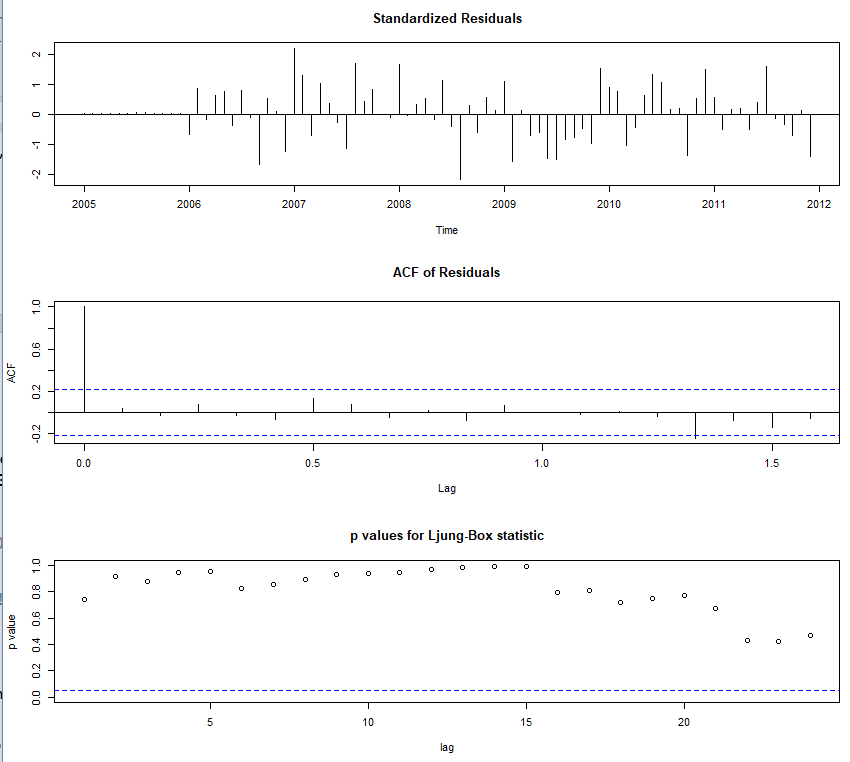
From the forecast plot, this model does a better job of forecasting than the previous models. The confidence intervals are narrower than the other forecast plots from visual inspection. The out-of-sample fit and bias are also better with this model, though not very good. The MAPE of this model is 1.98, which is the lowest of all the models, and the MASE is 0.73, which is also the lowest of all the models I ran. However, a 1.98 MAPE and a 0.73 MASE are not close to 0, and this indicates a still high forecast bias for the out-of-sample data.

1. (5 pts) Forecasting future monthly US electricity generation:
   * Use the Arima(…) function to fit the best model you have found thus far, run the model diagnostics to test the model validity and use it to extrapolate (forecast) the monthly generation of electricity in the US through the end of 2022.

The best model I found thus far was an ARIMA (2,0,3)(2,1,2) model with drift and no lambda transformation added. This model was found on the reduced training data set manually from question 7.

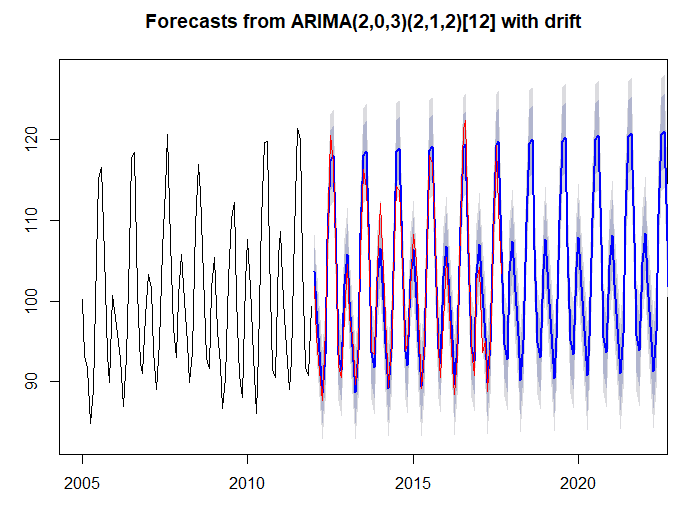


Model diagnostics to test the model validity:



The model diagnostics show that this model is valid, with the exception of one point on the ACF plot. All the standardized residuals are between -3 and +3. All of the ACF values of residuals are significant with the exception of one near lag = 1.3. All of the p-values are non-significant and not close to 0.

Forecast through the end of 2022:



Selection of some of the forecasted months through the end of 2022:

