University of California Santa Barbara

Final Project of:
PSTAT 174: Time Series

Professor:
Raya Feldman

Electricity Production of Time Series

Celeste Herrera
June 4, 2021

Table of Contents

I. Abstract

This Time Series project will elaborate on what factors contribute to the overall electric and gas production in the United States during the time frame of 1985 - 2018 with our frequency being considered to be under the monthly production output. The dataset used in this project was provided by Kaggle which more in depth was founded from Industrial production. My goal is to analyze and envision the following data of the electric and gas production with providing analysis and forecast of the industrial production. I explored the analytics of the data, there were numerous performances such as transforming , then following to the differencing, with the last application of diagnostics which acknowledged the use of the SARIMA which the model is best known for to be used with forecasting. When using the SARIMA model we were able to find a relatively close in year forecast which ranged from February 2017 to February 2018. I chose the model to be SARIMA(5, 1, 2) × $(2, 1, 2)_{12}$. It acts for the model of electric production and used it for year of electric production.

II. Introduction

As over time for many Americans the use electricity has increased over time it has been used for computer, cellphone, and even appliances. In the future more tools will be require electricity, therefore will need an influx of electricity production. With the data of the prior years we will be able to acknowledge how much will be needed in future endeavors.

The project will concentrate on factors that contribute to the overall electric and gas production in the United States during the time frame of 1985 - 2018. As stated previously the dataset will be provided from Kaggle. The objective then would be creating and establishing the prediction of the following 12 months from now where United States will be used for electric production.

When using Rstudio the code and packages will be acknowledged within the appendix. Throughout the datasets results many models will be tested will be acknowledged a method such as differencing and Box-Cox. Then running through the model we will take the step of forecasting. Then lastly when conducting all of the following it was safe to conclude that the model that was the best fit had a few points outside of the Confidence Interval.

III. Exploratory Data Analysis

A. Exploring the Data

The electricity production shows monthly values from the beginning to the period January 1985 to the end period of December 2018. It is observed their needs to be removal of the 12 points to help future predictions of forecasting in Figure 1.

Figure 1: Electric Production Time Series

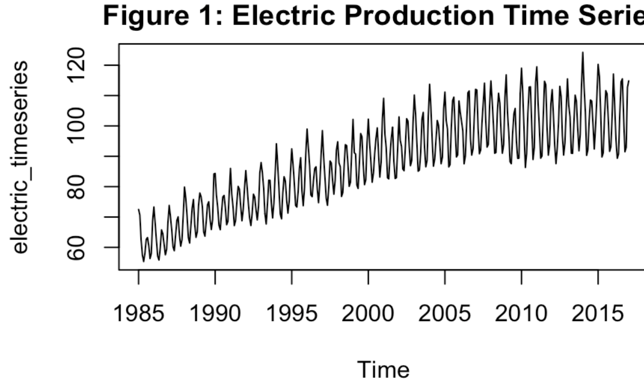
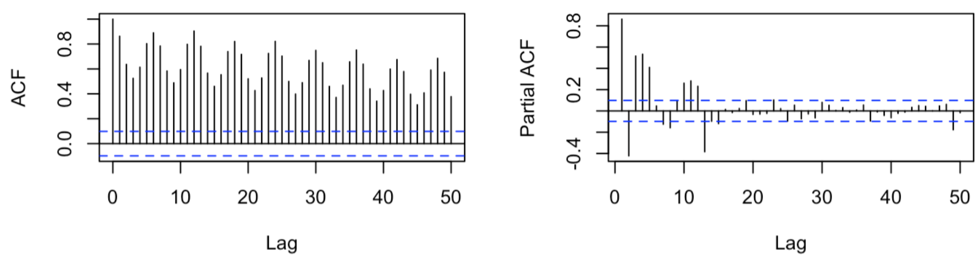


Figure 1 above demonstrates new amount of points which is 385 points. When we look for the variance of the new observations we get 233.83, although the variance is done before the transformation. In the Figure 1 we see that there is a trend from starting in the year of 1985 the graph tends to go upwards then seems to show a steady pace in an approximate year of 2010. There has been an increase of demand of electricity from now compared to the year of 1985.

Figure 2: ACF and PACF of Electric Production

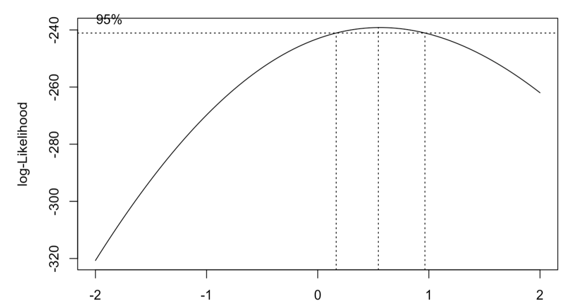


In Figure 2 we then use the function ACF (Auto Correlation Function) and PACF (Partial Auto Correlation Function) to better understand observations. In the ACF graph it has an overall trend of being large a seasonal component. The PACF graph should not be interpreted currently because of that reason differencing will be done to support the model estimate.

B. Box-Cox Transformation

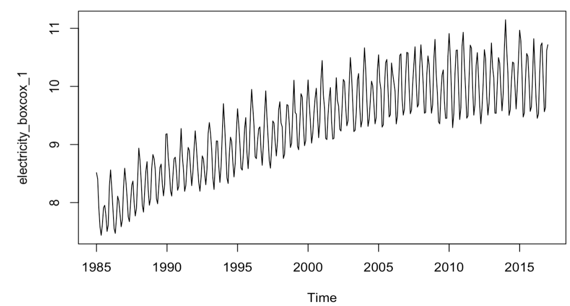
To stabilize the variance and try to make stationary one will explore the box-cox transformation. The reason for the box-cox is to minimize the given variance. When conducting the procedure we find for lambda to be 0.5454 which is shown in Figure 3.1.

Figure 3.1: Log-Likelihood of Box-Cox Transformation



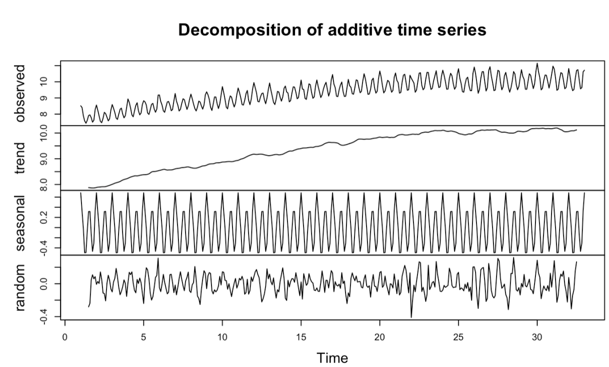
With the value of lambda given, transformation needs to be square root which is shown Figure 3.1. Since the difference it is acknowledged we want to see if the variance increase or decrease. The new variance is 0.6788 which will be shown in the Figure 3.2.

Figure 3.2: Box-Cox Transformed Data



We used few packages to transform the Box-Cox transformed data. Below will indicate specifics about it being transformed.

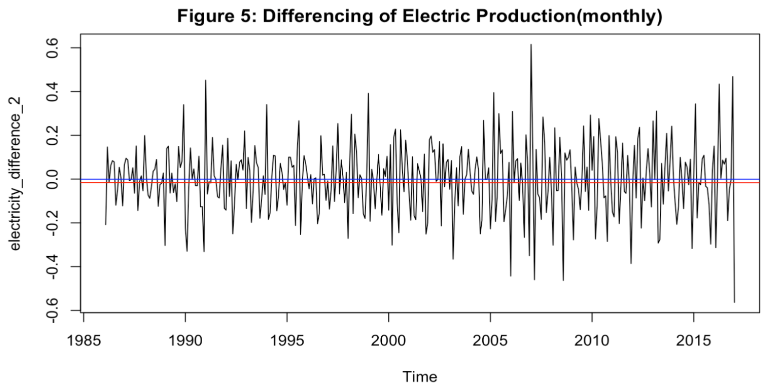
Figure 4: Transformed of Electric Production



In Figure 4 we will be transforming the data to be be decomposed. Therefore, we can state that the overall graphs of the random, seasonal, trend, and observed have positive upward trend. Although the best graph is to be the trend graph because it has a the most positive linear trend compared to the other graphs.

C. Differencing

In the differencing we tend to perform the logic rules off differencing which has a causation deseasonalizing our data and detrending. When deseasonalizing the data we will start at lag 12 because the seasonal component is at a value of 12 months. When looking Figure 4 there is common trend of a steady pattern therefore we need to remove the trend by detrending it by changing the lag to be 1.



Now in Figure 5 there is understanding of the Time Series model of Electric Production(Monthly). Overall we want variance to be smaller then the transformed data, retrieving result of 0.0259. In Figure 5 there are two lines indicated. The blue transformed data of the mean and the red line indicating that there is no type of trend and no type of seasonality in the graph.

IV. Model Building

Now acknowledging the ACF and PACF to create SARIMA(Seasonal Autoregressive Integrated Moving Average) to help develop our model that will best fit Electric Production dataset.

A. Analyzing ACF and PACF

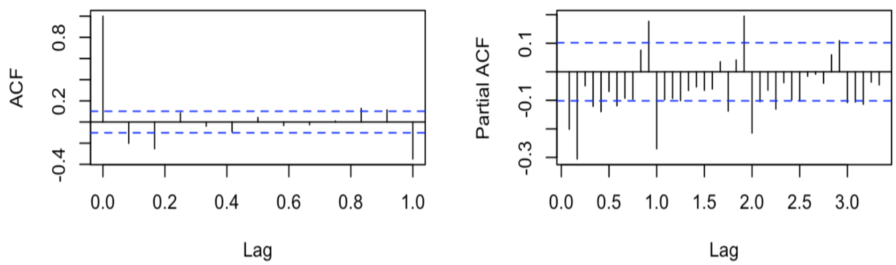
When looking at the image below one is able to indicate our coefficients with the following given:

p = Nonseasonal AR order	d = Nonseasonal differencing	q = Nonseasonal MA order
P = Seasonal AR order	D = Seasonal differencing	Q = Seasonal MA order
s = Seasonal Length		

p and q which are considered to be the nonseasonal components are found by lags of values through 1 and 11. P and Q are considered to be the seasonal components which is found after every 12 lags such as 12, 24, 36, and more. Then d and D received the value of 1 because Figure 5 showed the differencing.

In the ACF and the PACF I will be able to indicate the p and q letters which are considered to be the nonseasonal component. In the tests notice the difference in the size of the lags, to have a better look at the visibility of the graphs itself. At Lag 0 of the ACF and PACF (Figure 6) it show 0 is very minimalistic or has delay of a minute lag. Then continuing to look at the PACF of Figure 6 the first lag to be visibly see start value of 1. The letter p could be considered to be values of 1, 2, 4, 5, or 11. Although, we will not be considering value 11 because it is too large of a proper fit. Lastly, in Figure 6 when determining the letter q it could be the values of 0, 1, or 2.

Figure 6: ACF and PACF of Electric Production



Now will determine seasonal components of the letters P and Q. Looking for the values of P which can possibly be the value of 1 or 2 because there are small lags on the ends of the confidence intervals, which seems to be the value of 4 or 6. Therefore, I will not be looking at the lags due to the principle of parsimony. Q could be 1 or 2.

B. Selecting Model

After looking at the models below will be listed as the top 2:

Model 1: SARIMA(5, 1, 2) x $(2, 1, 2)_{12}$

Model 2: SARIMA(1, 1, 1) x $(2, 1, 2)_{12}$

After testing many models these four models previously shown to have the smallest AIC. When performing the AIC we found the values to be -528.3, -529.5, -529.1 and -527.9

The smallest AIC value -529.5 came from model 2 which is SARIMA(1, 1, 1) x $(2, 1, 2)_{12}$. Now to check the confidence intervals do not have a value of zero. Considering, zero is not in the confidence interval, will be eligible to continue the process with the model.

```
$fit

Call:
stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
    Q), period = S), include.mean = !no.constant, transform.pars = trans, fixed = fixed,
    optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:
         ar1      ma1    sar1     sar2     sma1    sma2
      0.5012  -0.9382  0.5438  -0.2529  -1.2780  0.4600
s.e.  0.0525   0.0205  0.2406   0.0671   0.2423  0.1964

sigma^2 estimated as 0.01304:  log likelihood = 271.77,  aic = -529.54

$degrees_of_freedom
[1] 366

$ttable
     Estimate     SE  t.value p.value
ar1    0.5012 0.0525   9.5449  0.0000
ma1   -0.9382 0.0205 -45.8226  0.0000
sar1   0.5438 0.2406   2.2600  0.0244
sar2  -0.2529 0.0671  -3.7658  0.0002
sma1  -1.2780 0.2423  -5.2747  0.0000
sma2   0.4600 0.1964   2.3424  0.0197

$AIC
[1] -1.382609

$AICc
[1] -1.382026

$BIC
[1] -1.310984
```

C. Model Diagnostics

Now checking model diagnostics, I will look at the residuals then follow with checking for normality then plot a graph. Below you will see the residuals of Figure 8.

Figure 7: Fitted Residuals, ACF and PACF of Electric Production

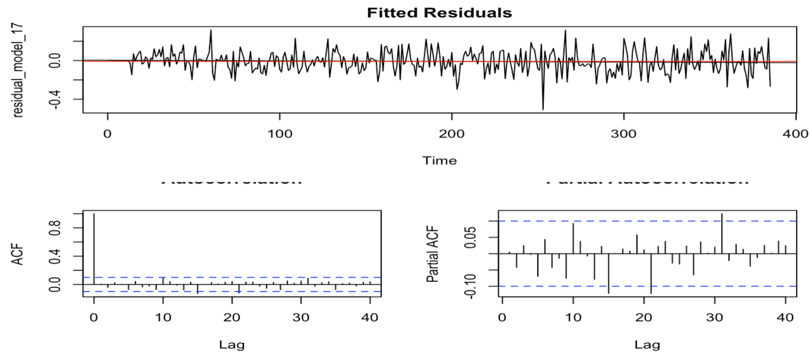


Figure 7 is within the confidence intervals. Although, few lags are outside of the interval. Then when found the mean is to be -0.008596415 and the variance is 0.01256123, therefore the model presents to be a good fit with a mean close to zero and a small variance.

```
        Box-Ljung test

data:  residual_model_17^2
X-squared = 34.757, df = 20, p-value = 0.02144

        Box-Ljung test

data:  residual_model_17
X-squared = 21.436, df = 13, p-value = 0.06473

        Box-Pierce test

data:  residual_model_17
X-squared = 20.688, df = 13, p-value = 0.07932
```

The p-values from the tests above we see that Box-Ljung test falls beneath the $\alpha = 0.05$. Although, the Shapiro Wilks test does pass with a given value of 0.0982 which is known be above the $\alpha = 0.05$. In this case, we will be looking at a new model to reevaluate.

D. Model Reselection

Moving on to the next smallest AIC which is -528.3 is considered to be Model 1. Model 1 is tested below and indicated as SARIMA(5, 1, 2) x $(2, 1, 2)_{12}$.

```
$fit

Call:
stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
    Q), period = S), include.mean = !no.constant, transform.pars = trans, fixed = fixed,
    optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:
          ar1     ar2     ar3     ar4      ar5      ma1      ma2    sar1     sar2     sma1   sma2
      -0.3905  0.4076  0.0014  0.0348  -0.0971  -0.0201  -0.8426  0.6199  -0.2519  -1.3813  0.547
s.e.   0.0715  0.0691  0.0657  0.0636   0.0574   0.0523   0.0495  0.1954   0.0696   0.1960  0.158

sigma^2 estimated as 0.01273:  log likelihood = 276.17,  aic = -528.34

$degrees_of_freedom
[1] 361

$ttable
     Estimate     SE  t.value p.value
ar1   -0.3905 0.0715  -5.4603  0.0000
ar2    0.4076 0.0691   5.8997  0.0000
ar3    0.0014 0.0657   0.0207  0.9835
ar4    0.0348 0.0636   0.5478  0.5841
ar5   -0.0971 0.0574  -1.6906  0.0918
ma1   -0.0201 0.0523  -0.3834  0.7016
ma2   -0.8426 0.0495 -17.0134  0.0000
sar1   0.6199 0.1954   3.1729  0.0016
sar2  -0.2519 0.0696  -3.6202  0.0003
sma1  -1.3813 0.1960  -7.0493  0.0000
sma2   0.5470 0.1580   3.4633  0.0006

$AIC
[1] -1.379476

$AICc
[1] -1.377618

$BIC
[1] -1.256691
```

Before looking at the AR values above, we need acknowledge to see if their is values 0 which need to be within the bounds of the coefficients. Coefficients of ar3, ar4 and ma1 all have the value 0. Since this is the case I need make minor changes to the model where the coefficients will have a set value of zero.

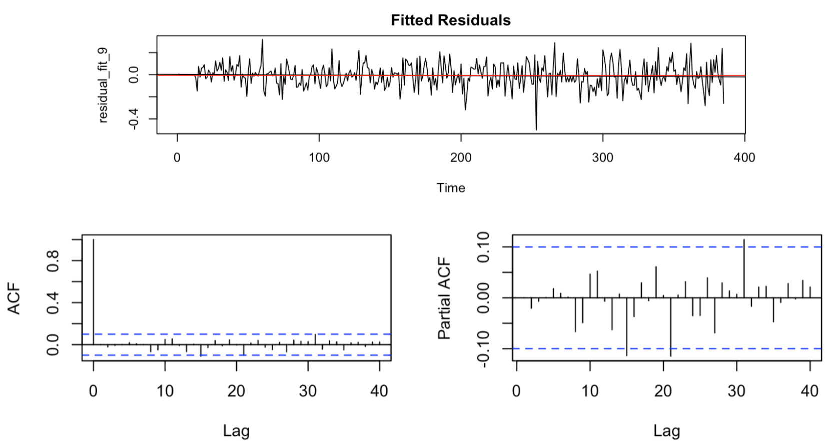
```
Call:
arima(x = electricity_boxcox, order = c(5, 1, 2), seasonal = list(order = c(2,
    1, 2), period = 12), transform.pars = FALSE, fixed = c(NA, NA, 0, 0, NA,
    0, NA, NA, NA, NA, NA), method = "ML")

Coefficients:
          ar1     ar2  ar3  ar4      ar5  ma1      ma2    sar1     sar2     sma1    sma2
      -0.4138  0.4130    0    0  -0.1119    0  -0.8433  0.6211  -0.2459  -1.3896  0.5511
s.e.   0.0488  0.0657    0    0   0.0477    0   0.0438  0.1928   0.0689   0.1941  0.1561

sigma^2 estimated as 0.01274:  log likelihood = 275.98,  aic = -533.95
```

The AIC of model 1 is smaller then the AIC of Model 2. Model 1 is a better fit than Model 2.

E. **Model Diagnostics**

Now looking at the model diagnostics again but for model 1, we look at the residuals then follow with checking for normality then plotting a graph. Figure 8 shows residuals of the fitted model.

Figure 8: Fitted Residuals, ACF and PACF of Electric Production

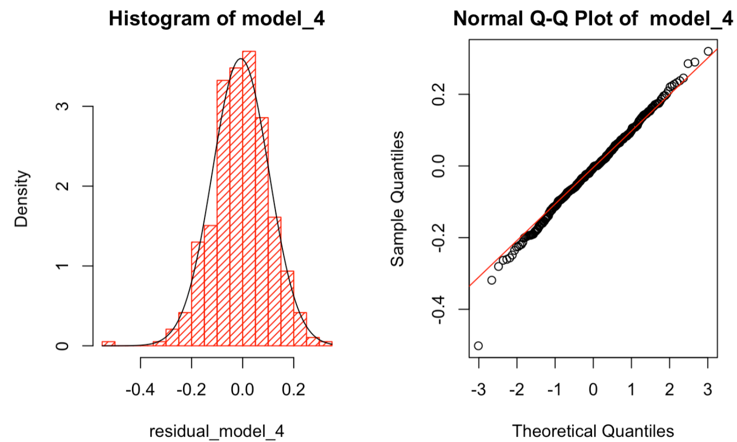


Looking at model 1(Figure 8) there are small lags outside of the confidence interval. Then new values of the mean is -0.007642 and 0.01229 for the variance. The mean and variance show model 1 is a better fit. Below is an indication of tests of whether it will also conclude if the model is a good fit.

```
        Box-Ljung test

data:  residual_model_4
X-squared = 13.685, df = 13, p-value = 0.3964

        Box-Pierce test

data:  residual_model_4
X-squared = 13.162, df = 13, p-value = 0.4354

        Box-Ljung test

data:  residual_model_4^2
X-squared = 27.869, df = 20, p-value = 0.1125
```

The results of the tests exceed $\alpha = 0.05$ in this case. As one of last tests I will be testing the Shapiro Wilks test to confirm that it is a good fit of a model. When looking at the p-value we get result of 0.122. It is known for it be above $\alpha = 0.05$ therefore the residuals of the model are normal. In Figure 9 will be able to visualize the detail of model 1.

Figure 9: Histogram and Normal Q-Q Plot



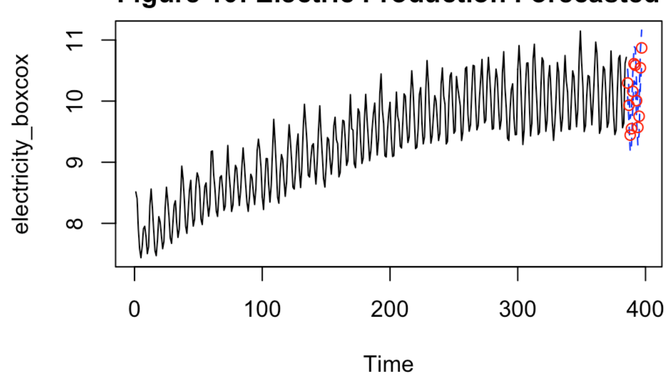
Looking Figure 9 we see the density curve and Normal Q-Q plot are normal which indicates that they passed the tests. With previous information I can state that the residuals of model 1 reflect Gaussian White Noise Distribution. Therefore, I can conclude that model 1 is the best fit for the dataset.

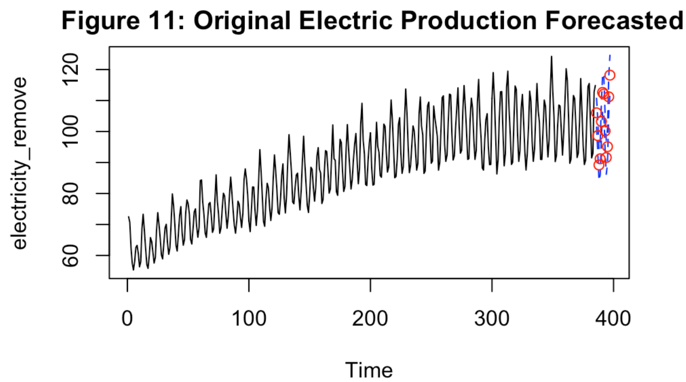
V. Forecasting

Model 1 was indicated as to be SARIMA(5, 1, 2) x (2, 1, 2)$_{12}$. The dataset forecasted for period of 12 months from the time of February 2017 to the time of January 2018 along with the known confidence interval of 95%.

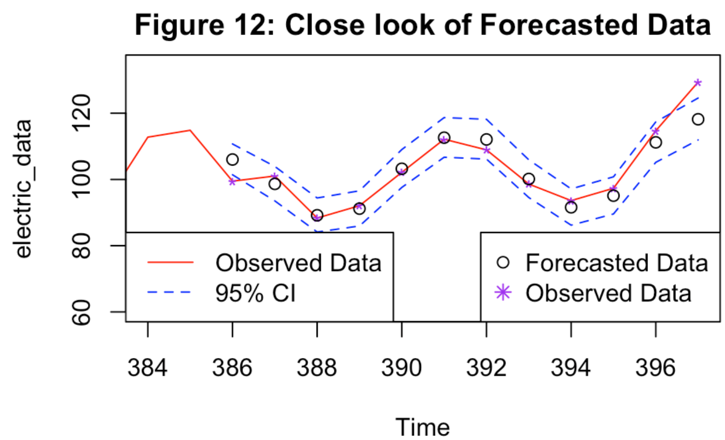
When looking at the original data compared to the transformed data we see similarities but have a better look with red points and the blue dashes.

Figure 10: Electric Production Forecasted





Looking at the Figure 10&11 there is better visualization of the transformed forecasted data and the original forecasted data.



In Figure 12 we see a close visual of the observed data and forecasted data with it also inside the range of the 95% confidence interval. The forecasted data points do much better then observed data points.

Upper <dbl>	Lower <dbl>	Observations <dbl>
101.42215	110.72111	99.4901
93.52954	103.92747	101.0396
84.12886	94.40143	88.3530
85.95088	96.58342	92.0805
97.59612	109.03279	102.1532
106.71405	118.68248	112.1538
106.15765	118.16637	108.9312
94.53963	105.92480	98.6154
86.19629	97.15376	93.6137
89.58139	100.79449	97.3359

The values of Figure 12 show that two points did not fall in the 95% confidence interval. There are many assumptions as to why that could possibly be but for now the assumption for it to be about the model not complying in the time period of February 2017 and January 2018. It is safe to state that Model 1: SARIMA(5, 1, 2) x $(2, 1, 2)_{12}$ is the best model for the dataset.

VI. Conclusion

In 12 month period of electric production the goal was to acknowledge the best fit for the model. When analyzing the data it was best found that SARIMA model was the best for the data. From the original model we were going to choose which was Model 2 which failed to comply with Box-Ljung test with having a small p-value. So we therefore, I complied with a new model. Taking account of Model 1 was the best solution because it had the second smallest AIC. In Model 1 we needed to acknowledge residuals and conduct tests to confirm that model 1 was a good fit. Model 1 was a good fit for the observed data. Although, when it came to forecasting we had two data points outside of the 95% confidence interval. In conclusion with the minor detail of the confidence interval we still found model 1 to be the best fit for the dataset.

VII. References

ShenbagaKumarS. (2018, September 19). Electricity Production. Retrieved May 15, 2021, from https://www.kaggle.com/shenba/electricity-production

Industrial Production: Utilities: Electric and Gas Utilities (NAICS = 2211,2). (2021, May 28). Retrieved May 30, 2021, from https://fred.stlouisfed.org/series/IPG2211A2N

Feldman, Raya. "PSTAT 174: Week 1-10: Lecture Notes." Time Series 174, 2 Jun. 2021, University of California, Santa Barbara .Microsoft PowerPoint presentation.

VIII. Appendix

```
# Retrieving directory for data
getwd()
setwd("/Users/celesteherrera/Documents/PSTAT174/Project")

electricity <- read.csv("Electric_Production.csv")
electric_data <- ts(electricity[,2])

# Remove unnecessary values for this moment that will be used
for forcasting and checking the variance.
electricity_remove<- electric_data[-c(386:397)]

var(electricity_remove)
# The variance of electricity_1 is 233.8271

# Now will reinsert the original values in this variable
electricity_new<- electric_data[386:397]

#Creating the plot for the start year of 1985 and also creating
the frequency
```

```
electric_timeseries <- ts(electricity_remove, start=c(1985,1,1),
frequency = 12)
ts.plot(electric_timeseries, main = "Figure 1: Electric
Production Time Series")

# Interpret graphs of ACF and PACF plots
par(mfrow = c(2,2))
acf(electric_data, lag.max = 50)
pacf(electric_data, lag.max = 50)

#Now will begin to perform a Box-Cox in order to find the
transformation

library(MASS)

transform = 1:length(electricity_remove)
fit_transform = lm(electricity_remove ~ transform)
boxcox_transform = boxcox(electricity_remove~transform, plotit =
TRUE)
lambda = boxcox_transform$x[which(boxcox_transform$y ==
max(boxcox_transform$y ))]
lambda
# lambda = 0.5454545

# Now to apply transformation,then follow with checking
variance, plot time series transformation

electricity_boxcox = sqrt(electricity_remove)
var(electricity_boxcox)
electricity_boxcox_1 <- ts(electricity_boxcox, start =
c(1985,1,1), frequency = 12)

ts.plot(electricity_boxcox_1)

# Begin decomposition of additive time series plot

library('ggplot2')
library('ggfortify')

value<- ts(as.ts(electricity_boxcox),frequency = 12)
decomposition <- decompose(value)
plot(decomposition)

# Transformed data of the difference and checking the variances

electricity_difference <- diff(electricity_boxcox_1,12)
var(electricity_difference)
```

```
# The variance of electricity_difference is 0.02949667

electricity_difference_2 <- diff(electricity_difference,1)
var(electricity_difference_2)
# The variance of electricity_difference is 0.02586812

# Plot the difference of the data for Electricity
Production(monthly)

ts.plot(electricity_difference_2, main = "Figure 5: Differencing
of Electric Production(monthly)")
 fitted_value <-
lm(electricity_difference_2~as.numeric(1:length(electricity_diff
erence_2)))

 abline(fitted_value, col = "red")
 abline(h = mean(electricity_difference_2), col = "blue")

# Plot the ACF and PACF for the differencing of the data for
Electricity Production(monthly)
par(mfrow = c(2,2))
acf(electricity_difference_2, lag.max = 12)
pacf(electricity_difference_2, lag.max = 40)

acf(electricity_difference_2, lag.max = 100)
pacf(electricity_difference_2, lag.max = 100)

# Goal is to find a SARIMA model with the lowest AIC
library(astsa)
model_1 <- sarima( xdata = electricity_boxcox, p = 5, d = 1, q =
0, P = 2, D = 1, Q = 2, S = 12, details = F)
model_1
model_2 <- sarima( xdata = electricity_boxcox, p = 5, d = 1, q =
0, P = 1, D = 1, Q = 2, S = 12, details = F)
model_2
model_3 <- sarima( xdata = electricity_boxcox, p = 5, d = 1, q =
1, P = 1, D = 1, Q = 2, S = 12, details = F)
model_3
model_4 <- sarima( xdata = electricity_boxcox, p = 5, d = 1, q =
2, P = 2, D = 1, Q = 2, S = 12, details = F)
model_4
model_5 <- sarima( xdata = electricity_boxcox, p = 4, d = 1, q =
0, P = 2, D = 1, Q = 2, S = 12, details = F)
model_5
model_6 <- sarima( xdata = electricity_boxcox, p = 4, d = 1, q =
0, P = 1, D = 1, Q = 2, S = 12, details = F)
model_6
```

```
model_7 <- sarima( xdata = electricity_boxcox, p = 4, d = 1, q =
0, P = 2, D = 1, Q = 1, S = 12, details = F)
model_7
model_8 <- sarima( xdata = electricity_boxcox, p = 4, d = 1, q =
0, P = 1, D = 1, Q = 1, S = 12, details = F)
model_8
model_9 <- sarima( xdata = electricity_boxcox, p = 4, d = 1, q =
1, P = 2, D = 1, Q = 2, S = 12, details = F)
model_9
model_10 <- sarima( xdata = electricity_boxcox, p = 4, d = 1, q
= 1, P = 1, D = 1, Q = 2, S = 12, details = F)
model_10
model_11 <- sarima( xdata = electricity_boxcox, p = 4, d = 1, q
= 1, P = 2, D = 1, Q = 1, S = 12, details = F)
model_11
model_12 <- sarima( xdata = electricity_boxcox, p = 4, d = 1, q
= 1, P = 1, D = 1, Q = 1, S = 12, details = F)
model_12
model_13 <- sarima( xdata = electricity_boxcox, p = 4, d = 1, q
= 2, P = 1, D = 1, Q = 1, S = 12, details = F)
model_13
model_14 <- sarima( xdata = electricity_boxcox, p = 2, d = 1, q
= 0, P = 2, D = 1, Q = 2, S = 12, details = F)
model_14
model_15 <- sarima( xdata = electricity_boxcox, p = 2, d = 1, q
= 0, P = 1, D = 1, Q = 2, S = 12, details = F)
model_15
model_16 <- sarima( xdata = electricity_boxcox, p = 1, d = 1, q
= 0, P = 1, D = 1, Q = 2, S = 12, details = F)
model_16
model_17 <- sarima( xdata = electricity_boxcox, p = 1, d = 1, q
= 1, P = 2, D = 1, Q = 2, S = 12, details = F)
model_17

# check residuals of model_41
residual_model_41<- residuals(model_41$fit)
mean(residual_model_41)
# the mean of residual_model_41 is -0.008596415
var(residual_model_41)
# the variance of residual_model_41 is 0.01256123

# plot the residuals for ACF and PACF of model_41
layout(matrix(c(1,1,2,3),2,2,byrow=T))
ts.plot(residual_model_41,main = "Fitted Residuals")
length_residual = 1:length(residual_model_41)
fitted_residual = lm(residual_model_41~length_residual)
abline(fitted_residual)
```

```
abline(h = mean(residual_model_41), col = "red")
acf(residual_model_41,main = "Autocorrelation", lag.max = 40)
pacf(residual_model_41,main = "Partial Autocorrelation", lag.max
= 40)

#Begin performance of box tests for model_41
Box.test(residual_model_41^2, lag = 20, type = c("Ljung-Box"),
fitdf = 0)

Box.test(residual_model_41, lag = 20, type = c("Ljung-Box"),
fitdf = 7)

Box.test(residual_model_41, lag = 20, type = c("Box-Pierce"),
fitdf = 7)

# Apply Shapiro test for model_41
shapiro.test(residual_model_41)

# Considering model_41 failed the box tests performed previously
we will now try a different model
# The new model will be model_9

adjusted_model_9 <- arima(electricity_boxcox,
order=c(5,1,2),seasonal=list(order=c(2,1,2),
period=12),fixed=c(NA,NA,0,0,NA,0, NA, NA, NA, NA, NA),
method="ML", transform.pars = FALSE)
adjusted_model_9

# Adjusted model_9 model
residual_model_9 <- residuals(adjusted_model_9)
mean_model_9 = mean(residual_model_9)
mean_model_9
# the mean of residual_model_9 is -0.007641781
variance_model_9 =var(residual_model_9)
variance_model_9
# the variance of residual_model_9 is 0.01228906

# Plot adjusted model’s residuals of ACF and PACF

layout(matrix(c(1,1,2,3),2,2,byrow=T))
ts.plot(residual_model_9,main = "Fitted Residuals")
t = 1:length(residual_model_9)
fit_residual_fit_9 = lm(residual_model_9~t)
abline(fit_residual_fit_9)
abline(h = mean(residual_model_9), col = "red")
acf(residual_model_9,main = "Autocorrelation", lag.max = 40)
```

```
pacf(residual_model_9,main = "Partial Autocorrelation", lag.max
= 40)

#Begin performance of model_9 for box tests
Box.test(residual_model_9, lag = 20, type = c("Ljung-Box"),
fitdf = 7)

Box.test(residual_model_9, lag = 20, type = c("Box-Pierce"),
fitdf = 7)

Box.test(residual_model_9^2, lag = 20, type = c("Ljung-Box"),
fitdf = 0)

# Apply Shapiro test for model_9
shapiro.test(residual_model_9)

# Since model_9 was a good test we will now presume to the plot
# Plot histogram and residuals
par(mfrow=c(1,2))
hist(residual_model_9,density = 20, breaks = 20, main =
"Histogram of model_9", col = 'red', prob = T)
curve( dnorm(x,mean_model_9,sqrt(variance_model_9)), add=TRUE )
qqnorm(residual_model_9, main = "Normal Q-Q Plot of  model_9")
qqline(residual_model_9,col ="red")

# Begin the procedure of forecasting
library(forecast)
adjusted_fit_9 <- arima(electricity_boxcox,
order=c(5,1,2),seasonal=list(order=c(2,1,2),
period=12),fixed=c(NA,NA,0,0,NA,0, NA, NA, NA, NA, NA),
method="ML")
forecast(adjusted_fit_9)

# Transforming data points
prediction_transformed <- predict(adjusted_fit_9, n.ahead = 12)
U_transformed= prediction_transformed$pred +
2*prediction_transformed$se
L_transformed= prediction_transformed$pred -
2*prediction_transformed$se
ts.plot(electricity_boxcox,
xlim=c(1,length(electricity_boxcox)+12), ylim =
c(min(electricity_boxcox),max(U_transformed)), main = "Figure
10: Electric Production Forecasted")
lines(U_transformed, col="blue", lty="dashed")
lines(L_transformed, col="blue", lty="dashed")
points((length(electricity_boxcox)+1):(length(electricity_boxcox
)+12), prediction_transformed$pred, col="red")
```

```
# Original Dataset with Forecast
prediction_original <- (prediction_transformed$pred)^2
U_transformed_2= (U_transformed)^2
L_transformed_2= (L_transformed)^2
ts.plot(electricity_remove,
xlim=c(1,length(electricity_remove)+12), ylim =
c(min(electricity_remove),max(U_transformed_2)), main = " Figure
11: Original Electric Production Forecasted")
lines(U_transformed_2, col="blue", lty="dashed")
lines(L_transformed_2, col="blue", lty="dashed")
points((length(electricity_remove)+1):(length(electricity_remove
)+12), prediction_original, col="red")

# Original Dataset with Forecast Zoomed In
ts.plot(electric_data,
xlim=c(length(electricity_remove)-1,length(electricity_remove)+1
2), ylim = c(60,max(U_transformed_2+10)), col="red", main =
"Figure 12: Close look of Forecasted Data")
lines(U_transformed_2, col="blue", lty="dashed")
lines(L_transformed_2, col="blue", lty="dashed")
points((length(electricity_remove)+1):(length(electricity_remove
)+12), electricity_new, pch = '*',col="purple")
points((length(electricity_remove)+1):(length(electricity_remove
)+12), prediction_original, col="black")
legend('bottomright', c('Forecasted Data', 'Observed Data'), bg
= 'white', pch = c(1,8), col = c('black', 'purple'))
legend('bottomleft', c('Observed Data', '95% CI'), bg = 'white',
lty = c(1,2), lwd = c(1,1), col = c('red', 'blue'))

# 95% CI with Actual Observation
observation<-data.frame(Lower=L_transformed_2,Upper=U_transforme
d_2,OBS=electricity_new)
names(observation)<-c("Upper","Lower","Observations")
observation
```