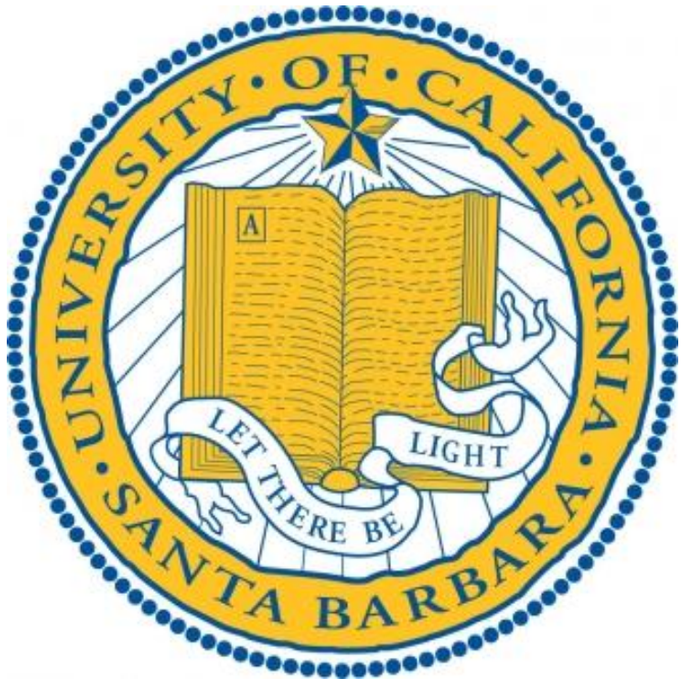


Natural Gas Consumption in the United States

Instructor: Raya Feldman

Author: Junyue Wang

Contents

Abstract

To understand the change in nationwide natural gas consumption, this project aims to build a time series model to interpret the data from Federal Reserve Economic Data and predict the future natural gas consumption. The model is trained on data from January 2000 - April 2019 and it is tested on May2019 - Feb 2020 data. After a proper transformation and analysis process, the final model is SARIMA (1, 1, 1) (0, 1, 1)_12. In the forecasting, all predictions lie within prediction intervals and are close to true values. Therefore, we can conclude that as long as no unexpected events happen, the sarima model is robust for forecasting.

Introduction

Natural gas is a fossil energy source that was formed deep beneath the earth's surface. It is one of the principal sources of energy for many of our day-to-day needs and activities. Gas consumption is directly related to social economic development as well as consumers' behaviors. In this project, I will explore the dataset *Natural Gas Consumption* from Federal Reserve Economic Data. It is released by U.S Transportation Data from the U.S. Bureau of Transportation Statistics. The data was collected monthly from Jan 2000 to Feb 2020 and there are 242 observations in total. The unit of gas consumption is in billion cubic feet. Our goal is to build up a powerful time series model which could be used to forecast the future gas consumption. To be specific with my own project, the goal is to predict the monthly gas consumption from 2019-05-01 to 2020-02-01.

Prior to the analysis, the dataset has been partitioned into a training set (232 observations) and a test set (the last ten observations). The training set is used to build the model and test set is used to validate the performance. If the predictions are accurate, we can say the model is robust and ready for application.

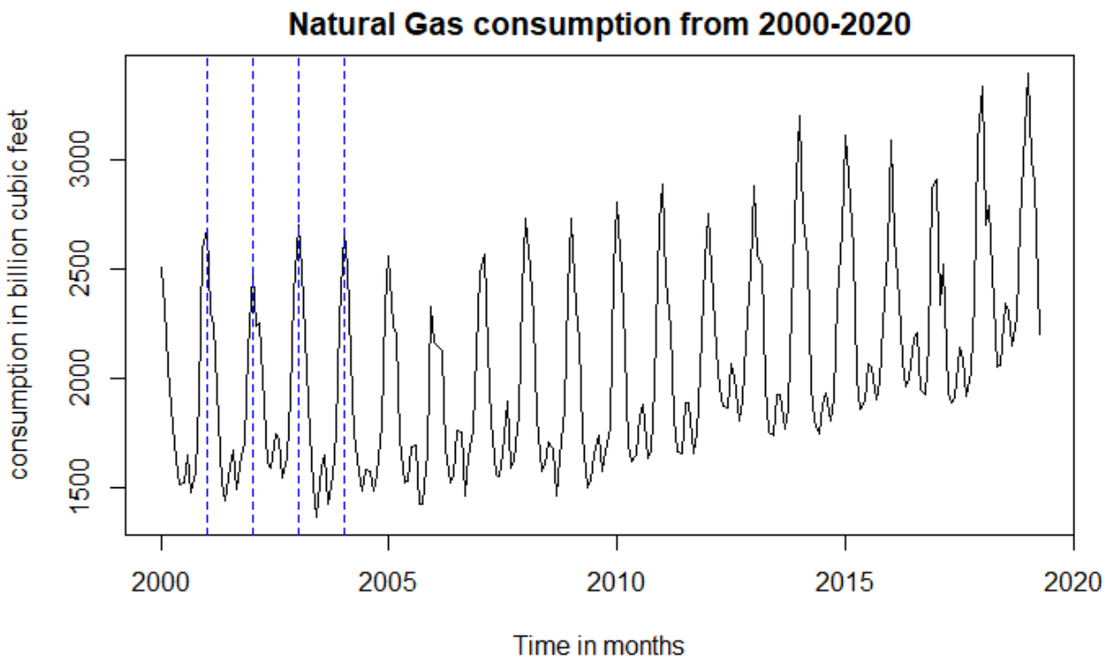
After converted into time series data, as expected, the raw data displays a strong seasonal pattern of consumption, indicating the existence of a certain consumption pattern and predictability. Besides, the increasing variation within a year of gas consumption suggests that the transformation is required to stabilize the variance and seasonal effect. From the histogram, the data is distributed with a long tail, therefore a box-cox transformation is also necessary to make data approximately normal. The upward trend suggests that a potential difference step is required to remove the trend.

To remove the seasonality and trend, the data was differenced respectively at lag 12 and at lag 1. Since an additional differencing at lag 2 to remove upward trend results in a larger variance, only one step differencing is adopted. Then according to acf and pacf plots, the range of the parameters can be estimated for candidate models. Next, I proceed with AICc and manage to find the top 4 models with the smallest AICc values. I choose the smallest SARIMA model to move on diagnostics checking and estimate coefficients with the MLE method. In the fitted residuals plot, an outlier was discovered in Jan 2001. After researching online, I found that according to the *Pipeline and Hazardous Materials Safety Administration (PHMSA)*, a United States Department of Transportation agency, 5 significant incidents with great impact were recorded in January 2001. They took place across the state, including Mississippi, Missouri, Utah and Kansas. Those accidents led to a huge amount of natural gas leaking and countless property damage. Apparently, those accidents could affect natural gas consumption in many ways and bring an unwanted result to our analysis. Therefore I decided to move forward with the fifteenth observations. The rest of the residuals perform well and pretty much resemble a white noise with mean zero and variance 1.922078e-10 .

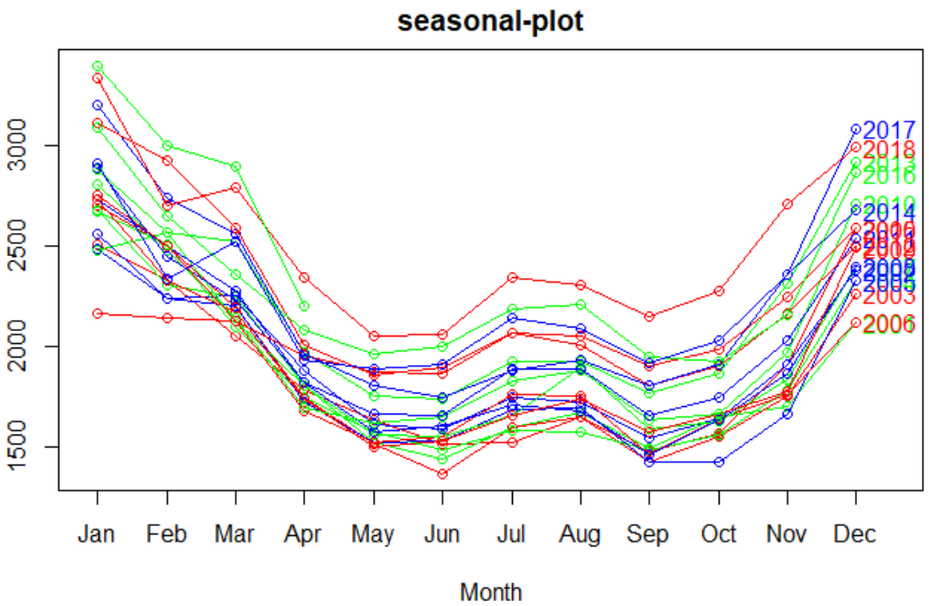
The final model is determined as SARIMA(1,1,1)*(0,1,1) with period 12.

Data Analysis

The data is retrieved from Federal Reserve Economic Data for Natural Gas consumption from Jan 2000 to Feb 2020. In total 242 monthly gas consumption data was recorded and provided for analysis. After I partitioned the plot from 2001 to 2004, a clear within year pattern is displayed. Two peaks appear at the beginning/end of a year and the consumption drops significantly in the time between.

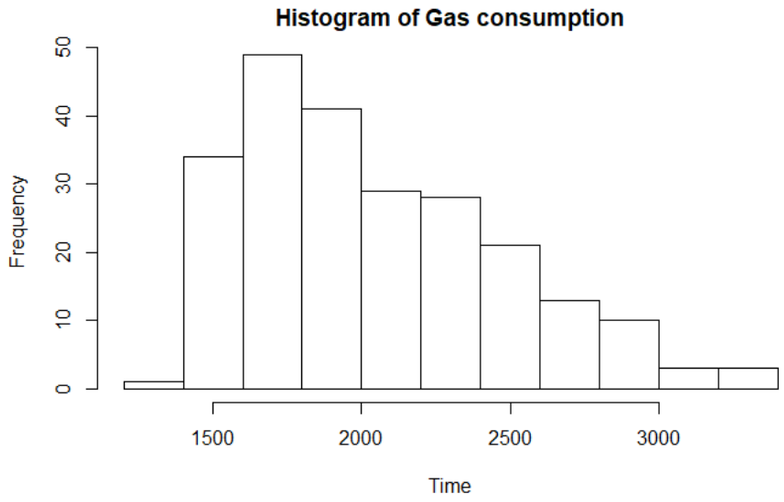


To clearly understand the monthly difference, the seasonal plot offers a better visualization. The pattern is strongly season related and people tend to consume more gas during cold months while less during warm months. It makes sense, as we know, a primary usage of natural gas includes the heating system. And natural gas was also used to generate electricity to support cooling AC in July or August.

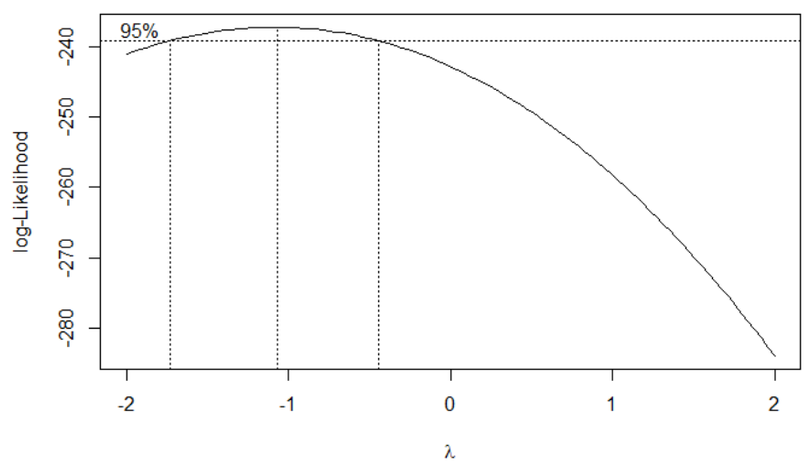


Data Transformation

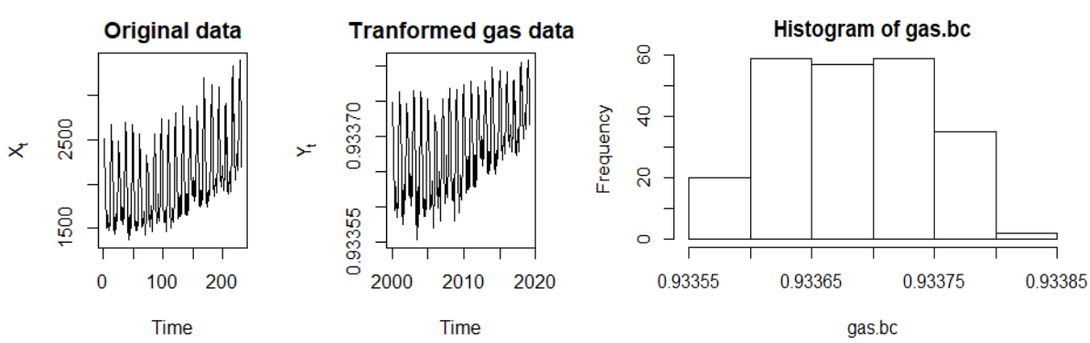
This histogram plot shows a heavy tail and therefore, I apply the box-cox transformation to approach a normally distributed data.



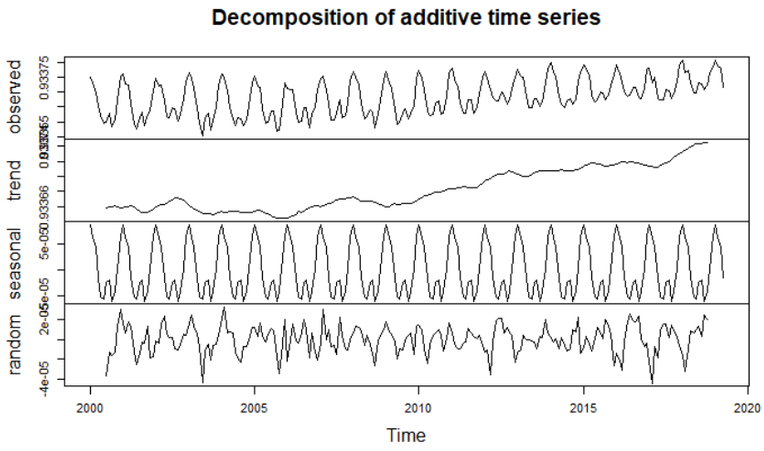
The box-cox transformation indicates a lambda value of -1.07.



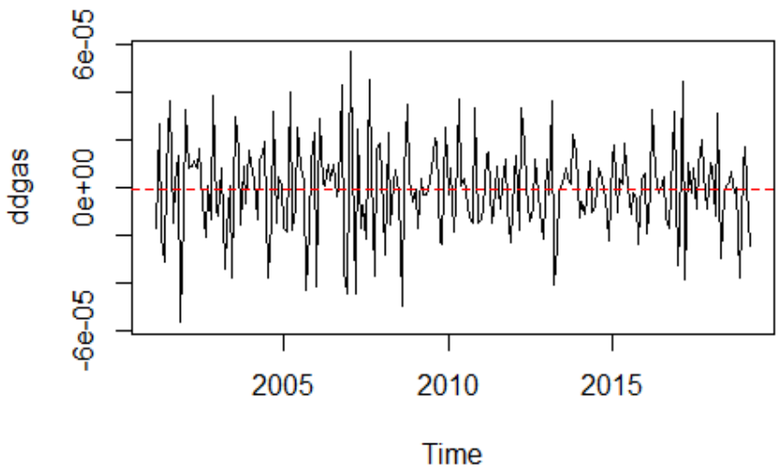
Plug in the value according to formula $f_\lambda = \lambda^{-1}(U_t^\lambda - 1)$, the transformed data provides a smaller variance (3.553669e-09) and normal data.



Now we move on to remove the trend and seasonality indicated above. A decomposition plot clearly shows seasonality and an almost linear trend preparing for difference.

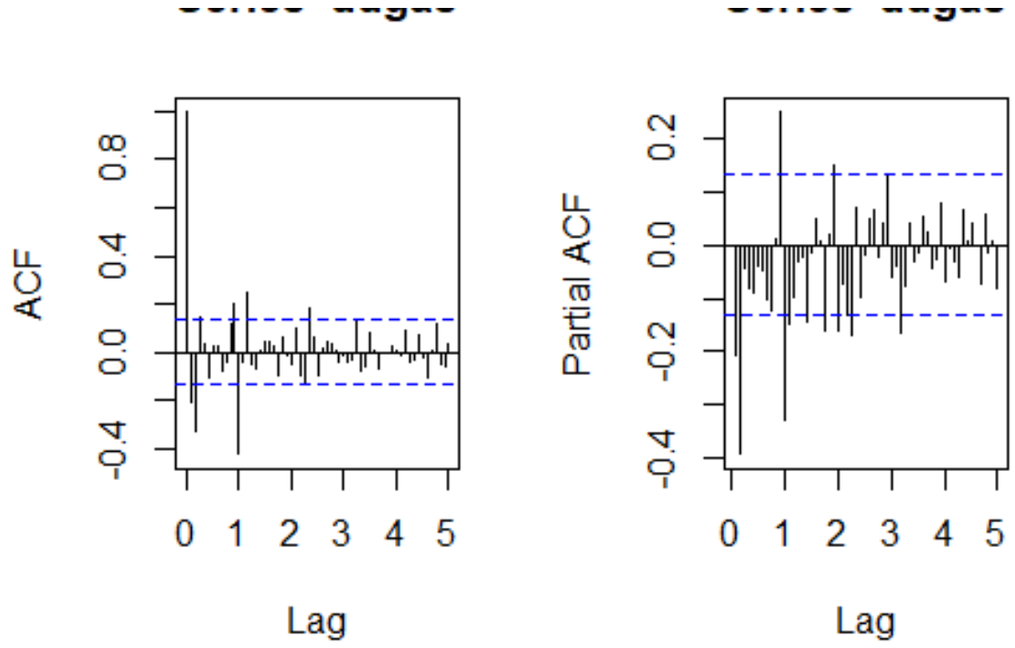


As a difference at lag 2 generates higher variance (over-differencing), I decide to keep one difference step for removing the upward trend. The data is collected monthly, and it displays the same pattern every year, so a difference at lag=12 is applied to remove the seasonality. The plot shows a stationary data with no trend and no seasonality, with mean -1.4e-07 and variance 3.7e-10. The data is ready for further analysis.



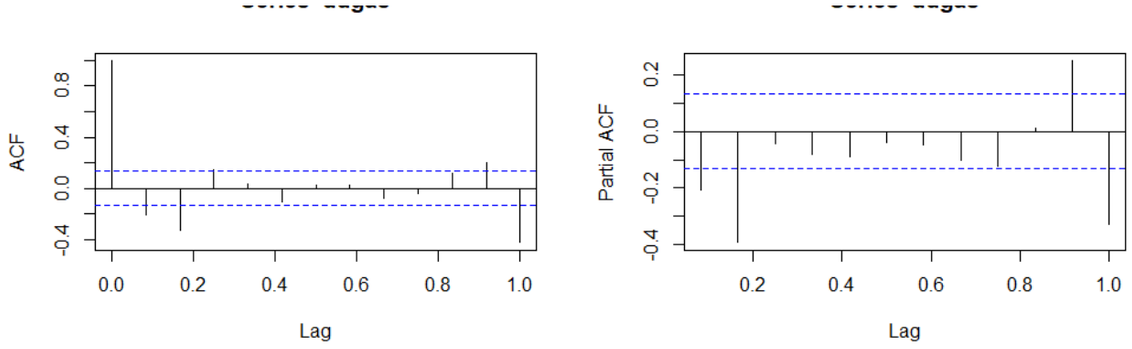
Preliminary model identification

Let Y_t denote the series ddgas and X_t denote the time series as X_t . Then we now have $Y_t = (1-B)(1-B^{12})X_t$. We now perform the Model modification part by using ACF and PACF to select possible model candidates for SARIMA model. First we look at the seasonal part acf\pacf plot.



The ACF shows a strong peak at h = 1s, A good choice for the MA part could be Q = 1. The PACF seems to be tailing off. Or perhaps cuts off at lag. So we may choose P=0, P=1 or P=2.

Modeling the non-seasonal part (p , d, q): In this case focus on the within season lags, h = 1,. . . ,11.



Although we have large values in lag 11 or lag 12, since large pacf lags tend to be affected by seasonal models, we choose smaller values p = 2. For acf, lag 3 slightly exceeds the confidence interval, therefore we will have options of q = from 1 to 3. The final model tends to include small parameters for model simplicity as well.

Next, I checked the candidate models with the AICc matrix. In an effort to find the best models, I choose all parameters from 0 to 2 or 0 to 3 for full consideration.

```
     p q P Q      AICc
[1,] 1 1 0 1 -4258.738
[2,] 2 1 0 1 -4257.272
[3,] 1 1 1 1 -4256.734
[4,] 2 1 1 1 -4255.236
```

For some values, the hessian metrics do not convergent and therefore I picked a few more candidates for testing. After trying to output $SARIMA(1,1,2)*(0,1,1)_{12}$,
I found it has the second smallest AICc value of $[1]-4257.876$

Here, I choose to use these two combinations as my candidate models.

$$Model1: SARIAM(1,1,1)*(0,1,1)_{12}$$
$$Model2: SARIMA(1,1,2)*(0,1,1)_{12}$$

AICc values are not enough to determine the model. Next, I will conduct diagnostic checking and see if one of them fails the residual assumptions including normality, independence and constant variance of errors.

Diagnostic Checking

To ensure two models are stationary and invertible, I plotted the root on a unit circle. (Figure 1 in appendix) All roots lie outside the unit circle and therefore we conclude both models are stationary and convertible.

Applying the coefficients, we expand two models as follows :

$(1 - 0.5186B)(1 - B)(1 - B^{12})X_t = (1 - 0.9477B)(1 - 0.7227B^{12})Z_t$

where $Z_t \sim (0, 1.915e^{-10})$

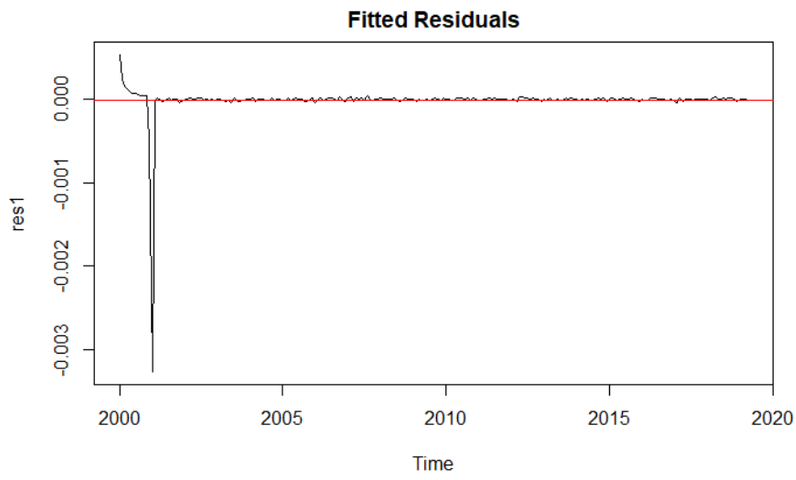
$(1-0.3451B)(1-B)(1-B^{12})X_t = (1-0.726B-0,1944B^2)(1-0.7215B^{12})Z_t$

Where $Z_t \sim (0, 1.924e^{-10})$

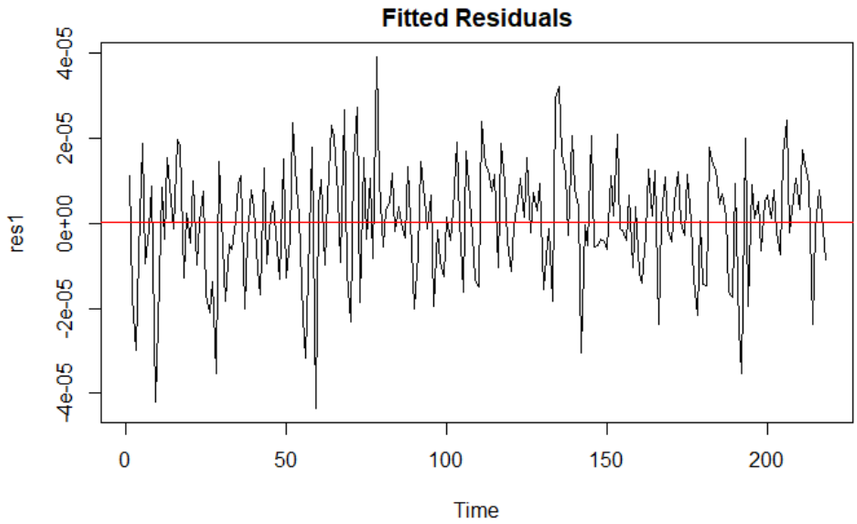
However, from the output coefficients, I suspect that model2 may not be appropriate in some sense because the confidence interval of coefficients for AR and MA2 part include 0, which means they may not be necessary.If MA2 is unnecessary, the model2 could be simplified to model 1.

```
Call:
arima(x = gas.bc, order = c(1, 1, 1), seasonal = list(order = c(0, 1, 1), period
= 12),
    method = c("ML"))

Coefficients:
         ar1      ma1     sma1
      0.5186  -0.9477  -0.7227
s.e.  0.0672   0.0278   0.0499

sigma^2 estimated as 1.924e-10:  log likelihood = 2133.42,  aic = -4258.84


Call:
arima(x = gas.bc, order = c(1, 1, 2), seasonal = list(order = c(0, 1, 1), period
= 12),
    method = c("ML"))

Coefficients:
         ar1     ma1      ma2     sma1
      0.3451  -0.726  -0.1944  -0.7215
s.e.  0.1791   0.195   0.1639   0.0499

sigma^2 estimated as 1.915e-10:  log likelihood = 2134.03,  aic = -4258.05
```

The fitted residuals plot indicates an outlier for both models in Jan 2001. After searching online, there are 5 major natural gas accidents in Jan 2001 and they may directly affect natural gas consumption. Therefore I started from March 2001 residuals to avoid the huge impact of that data point.



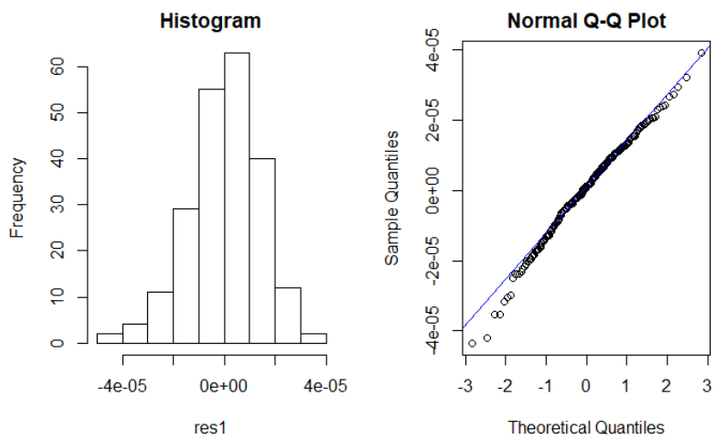
Here, we can see that the residuals perform very well after dropping the Jan 2001 data point, and ACF plot has all lags within the confidence interval while pacf has only one lag slightly exceeds the confidence interval. Model 1 and Model 2 share approximately similar fitted residuals plot. From the plot below, y-axis values are small and we can conclude the residuals have constant variance.



Next, I conducted a Box-Pierce test, Ljung-Box test for linear dependence check. A Mcleod-Li test was used for a non-linear dependence check.
From the result, both models pass the Box-pierce test and Mcleod-Li test. But p-values for the Ljung-Box test are 0.04608 for model1 and 0.04729 for model2, indicating that both models almost pass the Ljung-Box test. However, since the p-values are very close to 0.05 and the residuals plot resemble a white noise very much, I would stick to box-pierce test results and conclude the residuals are closely white noise. (R output in appendix 2) Hence we do not reject the assumption of uncorrelated residuals.

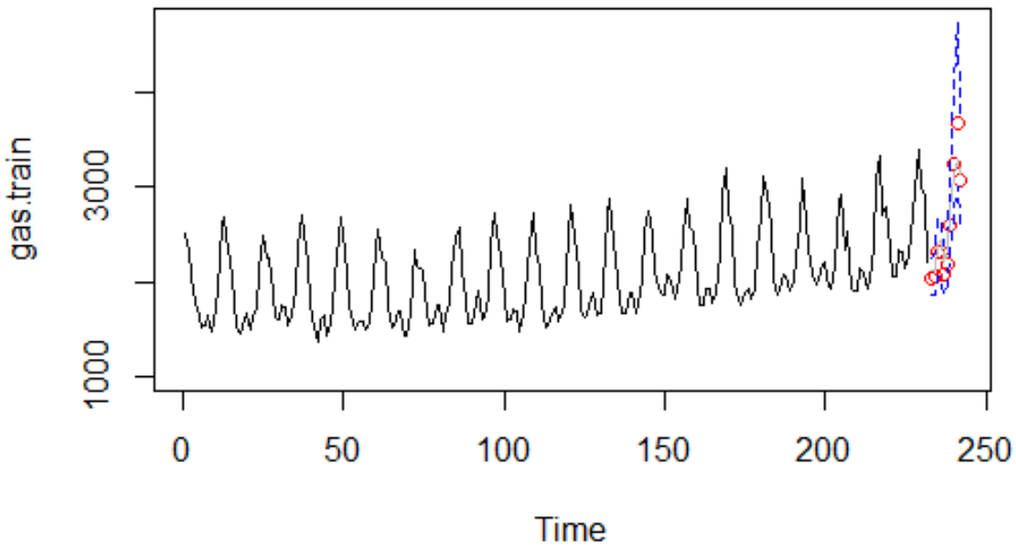
Finally, I choose to keep Model 1 with smaller AICc value $SARIMA(1,1,1) * (0,1,1)_{12}$ as my final model. I dropped Model 2 because of model simplicity and it can be reduced to Model1 as the extra MA2 process is determined to be unnecessary. (confidence interval of coefficients contains zero for MA2)

To ensure we have the correct prediction interval later, now I check whether residuals are normally distributed. The histogram and Q-Q norm of residuals show that they are normal. Further, shapiro-test provides p-value of 0.18, which means we fail to reject the null hypothesis that residuals are normal.

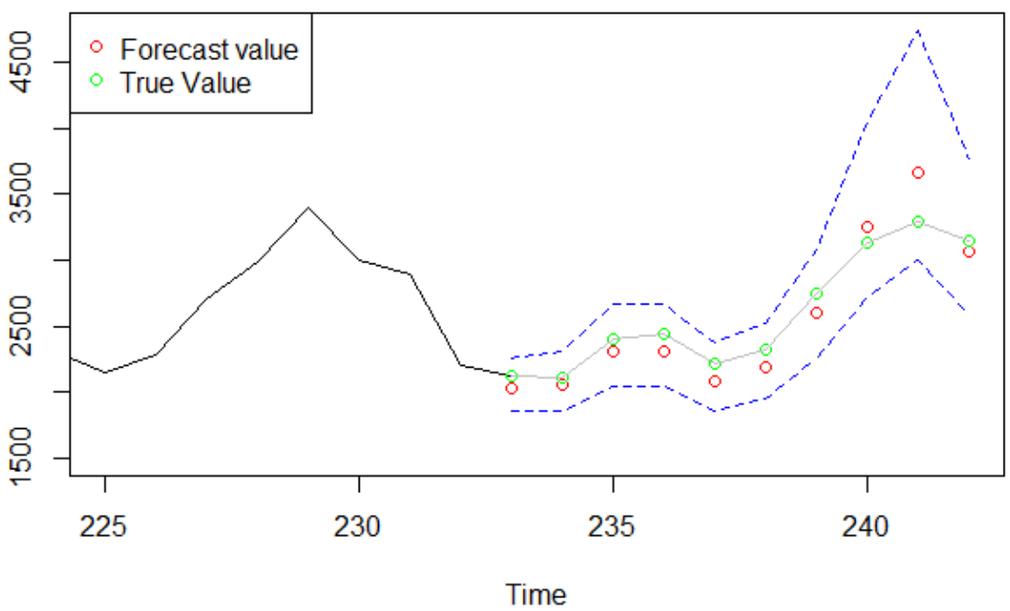


Data Forecast

The last step is to forecast the gas consumption from May2019 - Feb2020 with final model $SARIMA(1,1,1)(0,1,1)_{12}$. The forecast based on the original data is in the graph below.



The zoomed picture shows clearly that all predictions are within prediction intervals and get very close to true values. This demonstrates that our model is robust and predictions are accurate.



Conclusion

The project aims to build a time series model for gas consumption forecast. The data is splitted into a training set for model building and a test set(last 10 observations) for performance evaluation.

After doing preliminary analysis, I found the object is closely related with time and seasons.Box-cox transformation is applied to data for normality and smaller variance. Then I get to remove the trend and seasonality by differencing at lag1 and lag12. In this case, the parameters are determined as S=12,D=1,d=1. With the assistance of ACF\PACF plot and AICc values, two candidate models are ready to proceed with diagnostics checking. They both pass the Box-pierce test, Mcleod-Li test and almost pass the Ljung-Box test, however, I would choose my final model as SARIMA(1,1,1)(0,1,1)_12 out of model simplicity.

The expanded formula including coefficients :
$SARIMA(1,1,1)(0,1,1)_{12}$

$$(1-0.5186B)(1-B)(1-B^{12})X_t = (1-0.9477B)(1-0.7227B^{12})Z_t$$

where $Z_t \sim (0, 1.915e^{-10})$

According to acf\pacf and fitted residuals plot, Its residuals are quite normal and stationary after taking off the outliers. Therefore it is valid for forecasting. The predictions all lie within prediction intervals and the values are close to true values. Although the model has great prediction power, it is well worth noticing that some unusual events may affect the results. For the first quarter of 2020, people have been impacted by COVID-19 and unexpected influences may take place for natural gas consumption.

The completeness of this project is attributed to the assistance of Professor Raya Feldman, TA Sunpeng Duan and Nicole Yang. Thanks to all of you helping me with the project.

APPENDIX

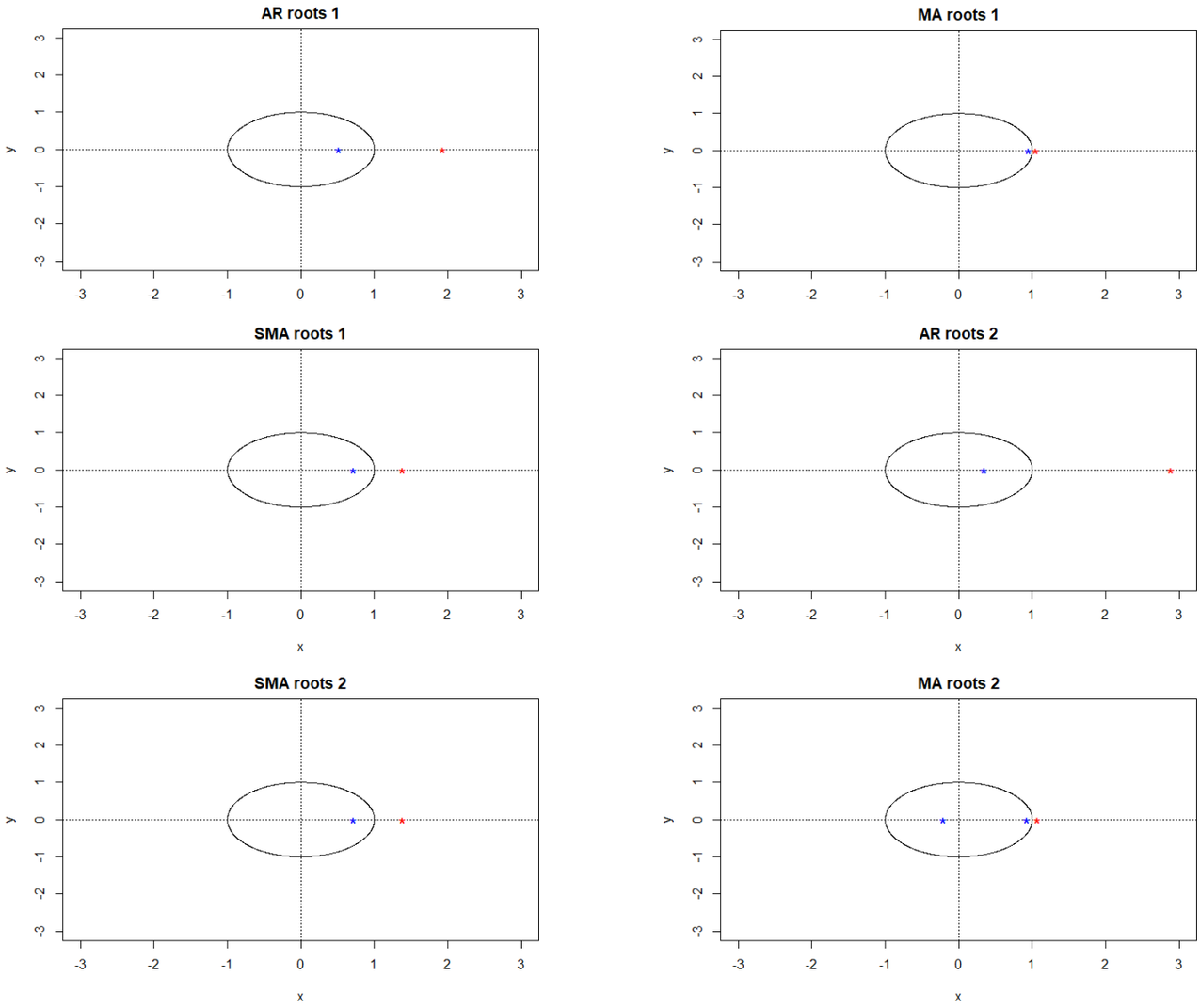


Figure 1

```{r}
Box.test(res1, lag = 15, type = c("Box-Pierce"), fitdf = 3)
Box.test(res1, lag = 15, type = c("Ljung-Box"), fitdf = 3)
Box.test(res1^2, lag = 15, type = c("Ljung-Box"), fitdf = 0)
```

```
        Box-Pierce test

data:  res1
X-squared = 20.255, df = 12, p-value = 0.06242


        Box-Ljung test

data:  res1
X-squared = 21.306, df = 12, p-value = 0.04608


        Box-Ljung test

data:  res1^2
X-squared = 7.4552, df = 15, p-value = 0.9438
```

```{r}
Box.test(res2,lag=15,type=c("Box-Pierce"),fitdf = 4)
Box.test(res2,lag=15,type=c("Ljung-Box"),fitdf = 4)
Box.test(res2^2, lag = 15, type = c("Ljung-Box"), fitdf = 0)
```

```
        Box-Pierce test

data:  res2
X-squared = 18.819, df = 11, p-value = 0.06443


        Box-Ljung test

data:  res2
X-squared = 19.86, df = 11, p-value = 0.04729


        Box-Ljung test

data:  res2^2
X-squared = 6.6401, df = 15, p-value = 0.9669
```

Figure 2

```{r,warning=FALSE}
library ( astsa )
library ( tseries )
library (MASS)
library ( forecast )
library (GeneCycle)
library ( ggplot2 )
library(ggfortify)
```

```{r}
setwd(dir = "C:/Users/21ans/Desktop/174/project")
gasorg <- read.csv("NATURALGAS.csv")
#View(gasorg)
gas.train = gasorg[1:232,2]

gas.test = gasorg[233:242,2]

```

Introduction

This dataset is collected monthly on domestic Natural Gas Consumption. The unit is Billion Cubic Feet.

Natural gas is a fossil energy source that formed deep beneath the earth's surface. Natural gas contains many different compounds. Gas consumption is directly related to many environmental issues as well as consumer's behaviors. Therefore, it is useful to understand the gas consumption pattern In the project, we will build a powerful model to fully explain the consumption pattern and forecast the future seasonal consumption.

```{r}
gas <- ts(gas.train,frequency = 12, start = c(2000,1)) #Monthly data points
ts.plot(gas,main = "Natural Gas consumption from 2000-2020", ylab = "consumption in billion cubic feet",xlab = "Time in months")
abline(v = ts(c(2001,2002,2003,2004)), col = "blue", lty = 2)
```

Frome the graph, we conclude that Gas data is not stationary because of the upward trend and seasonality.

```{r}
seasonplot(gas,12,col=rainbow(3),year.labels =TRUE,main="seasonal-plot")
```

```{r}
decomp <- decompose(gas)
plot(decomp)
```

The seasonal plot shows a clear graph of variations with seasons. Gas was consumed low amount during summer and high amount during winter or cold seasons. This corresponds to our guess and reality.

```{r}
hist(gas.train,main = "Histogram of Gas consumption",xlab = "Time")
```
The histogram plot does not show a relatively symmetric pattern, hence a transformation is probably necessary as the variance changes with time.

```{r}
library(MASS)
```

```
t = 1:length(gas)
fit = lm(gas ~ t)
bcTransform = boxcox(gas ~ t,plotit = TRUE)
bcTransform$x[which(bcTransform$y == max(bcTransform$y))]
```

Since 0 and 1 are not in confidence interval, we want to use box-cox transformation with maximized lambda value applied.

```{r}
lambda = bcTransform$x[which(bcTransform$y == max(bcTransform$y))]
gas.bc = (1/lambda)*(gas^lambda-1)
op <- par(mfrow = c(1,2))
ts.plot(gas.train,main = "Original data",ylab = expression(X[t]))
ts.plot(gas.bc,main = "Tranformed gas data", ylab = expression(Y[t]))
var(gas.bc);var(gas.train)
```

$ f_{\lambda} = \lambda^{-1}(U_{t}^{\lambda} - 1)$

We can see the box-cox transformation greatly stablizes the variance for gas data. However, the increasing trend still exists, meaning the data is not yet stable for analysis. In acf plot below, we can also observe the seasonality.

```{r}
hist(gas.bc)
```

```{r}
decomp <- decompose(gas.bc)
plot(decomp)
```

Now we proceed to remove the trend and seasonality.

```{r}
dgas <- diff(gas.bc, 12) # De-seasonalize
ddgas <- diff(dgas,1) # De-trend
ts.plot(ddgas)
abline(h = mean(ddgas),lty = 2,col = "red")
mean(ddgas);var(ddgas)
```

Then we check two steps differencing by comparing the variance and to decide on a number of differences.

```{r}
y1 = diff(ddgas, 1)
var(ddgas);var(y1)
```

Since the variance increased for the second difference, this step is unnecessary and I decided to difference the data only once. Therefore D=1 and d=1. The data was collected monthly, so the seasonal part would naturally be 12.

Let Y_t denote the series ddgas and X_t denote the time series as X_t . Then $Y_t = (1-B)(1-B^{12})X_t$ We now perform the Model modification part by using ACF and PACF to select possible model candidates.

```{r}
acf(ddgas, 100)
pacf(ddgas, 100)
```

SARIMA(p = 2,1,q = 2 or 3) (P = 1 or 2,1, Q= 1)

```{r}
acf(ddgas, 12)
pacf(ddgas, 12)
```

Now, we want to model Y_t with the sarima model. We applied one seasonal differencing so D =1 at lag s =12.

The ACF shows a strong peak at h = 1s, A good choice for the MA part could be Q = 1. In pacf plot, lag 2 slightly exceeds confidence interval, so I will test P = 1 or P = 2 later with AICc. For the non-seasonal part, we observe the within year acf and pacf plot. Although we have large values in lag 11 or lag 12, since pacf tends to overestimate p, we choose smaller values p = 2. For acf, lag 3 slightly exceeds the confidence interval, therefore we will have options of q = 1-3 and test later.

```{r, warning=FALSE}
library(astsa)
library(qpcR)
aiccs = matrix(NA, nr = 54, nc = 5)
colnames(aiccs) = c("p", "q","P","Q","AICc")
i = 0
for (Q in c(1)){
  for (q in c(0,1)){
    for(P in c(1)){
      for(p in c(0:2)){
        aiccs[i+1,1] = p
        aiccs[i+1,2] = q
        aiccs[i+1,3] = P
        aiccs[i+1,4] = Q
        aiccs[i+1,5] = AICc(arima(gas.bc,order = c(p,1,q),seasonal=list(order = c(P,1,Q),period = 12),method = c("ML")))
        print(c(p,q,P,Q))
        i = i+1
  }
    }
  }
}
aiccs[order(aiccs[,5])[1:4],]

```

Since some combinations failed to generate AICc value, I run the individual function and compare the AICcs.
```{r}
AICc(arima(gas.bc,order = c(1,1,1),seasonal=list(order = c(0,1,1),period = 12),method = c("ML")))

AICc(arima(gas.bc,order = c(1,1,1),seasonal=list(order = c(1,1,1),period = 12),method = c("ML")))

AICc(arima(gas.bc,order = c(1,1,1),seasonal=list(order = c(0,1,2),period = 12),method = c("ML")))

AICc(arima(gas.bc,order = c(2,1,1),seasonal=list(order = c(0,1,1),period = 12),method = c("ML")))

AICc(arima(gas.bc,order = c(2,1,1),seasonal=list(order = c(1,1,1),period = 12),method = c("ML")))

AICc(arima(gas.bc,order = c(2,1,1),seasonal=list(order = c(2,1,2),period = 12),method = c("ML")))

AICc(arima(gas.bc,order = c(2,1,2),seasonal=list(order = c(0,1,2),period = 12),method = c("ML")))

AICc(arima(gas.bc,order = c(1,1,2),seasonal=list(order = c(0,1,1),period = 12),method = c("ML"))) # would work better

AICc(arima(gas.bc,order = c(1,1,1),seasonal=list(order = c(0,1,2),period = 12),method = c("ML")))

AICc(arima(gas.bc,order = c(2,1,2),seasonal=list(order = c(2,1,1),period = 12),method = c("ML")))
```

```

For some values, the hessian metrics do not convergent and therefore I picked a few candidates for testing. I found $SARIMA(1,1,1)*(0,1,1)_{12}$ may be a better fit since it is also supported within year acf and pacf plot. I will treat both candidates model with diagnostic checkings. If it fails somewhere, I can go back and check with other candidates.

```{r}
gas.fit1 = arima(gas.bc,order= c(1,1,1),seasonal = list(order = c(0,1,1),period=12),method = c("ML"))
gas.fit1
```

```{r}
gas.fit2 = arima(gas.bc,order = c(1,1,2),seasonal = list(order = c(0,1,1),period = 12),method = c("ML"))
gas.fit2
```

```{r}
source("plot.roots.R")
plot.roots(NULL,polyroot(c(1,-0.5186)),main = "AR roots 1")
plot.roots(NULL,polyroot(c(1,-0.9477)),main = "MA roots 1")
plot.roots(NULL,polyroot(c(1,-0.7227)),main = "SMA roots 1")
plot.roots(NULL,polyroot(c(1,-0.3451)),main = "AR roots 2")
plot.roots(NULL,polyroot(c(1,-0.726,-0.1944)),main = "MA roots 2")
plot.roots(NULL,polyroot(c(1,-0.7215)),main = "SMA roots 2")
polyroot(c(1,-0.726,-0.1944))
polyroot(c(1,-0.9477))
```

For the first possible candidate, the roots all lie outside the unit circle. And the absolute values for all the coefficients are less than 1. Thus, we conclude that model 1 and model 2 are both causal and invertible.

Diagonostics checking

Two best models had been identified to fit our data. Now, before moving to forecasting, we should check with the normality, independence and constant variance of residuals of our model.

```{r}
res1 = residuals(gas.fit1)
res1 = res1[c(15:length(res1))]
#layout(matrix(c(1,1,2,3),2,2,byrow=F))
ts.plot(res1,main = "Fitted Residuals")
abline(h = mean(res1), col = "red")
acf(res1,main = "Autocorrelation")
pacf(res1,main = "Partial Autocorrelation")
var(res1)
```

```{r}
res2 = residuals(gas.fit2)
res2 = res2[c(15:length(res2))]
#layout(matrix(c(1,1,2,3),2,2,byrow=F))
ts.plot(res2,main = "Fitted Residuals")
abline(h = mean(res2), col = "red")
```
```

```
acf(res2,main = "Autocorrelation")
pacf(res2,main = "Partial Autocorrelation")

```

The fitted residuals plot indicates the residuals are stationary and resemble white noise. The result shows that most of the values lie within the bound (denoted by the blue dotted lines). For some that exceed the bound, they can be seen as the outliers in our dataset. However, in my fitted residuals plot, the variance looks weird in the beginning. There was a sharp change in January 2001, so I tried to move the residuals ahead by 2 months (that means I will start from March 2001).

```{r}
Box.test(res1, lag = 15, type = c("Box-Pierce"), fitdf = 3)
Box.test(res1, lag = 15, type = c("Ljung-Box"), fitdf = 3)
Box.test(res1^2, lag = 15, type = c("Ljung-Box"), fitdf = 0)
```

```{r}
Box.test(res2,lag=15,type=c("Box-Pierce"),fitdf = 4)
Box.test(res2,lag=15,type=c("Ljung-Box"),fitdf = 4)
Box.test(res2^2, lag = 15, type = c("Ljung-Box"), fitdf = 0)
```

According to Box-pierece Test and Box-Ljung test, we fail the reject the white noise hypothesis becasue P-value is greater than 0.05. From p-value of McLeod-Li test, the residuals are independent.

Next, we probably want to know whether the residuals are normally distributed. The reason for this step is to determine our prediction interval later. If the residuals do not follow normal distribution, we will consider a change of coefficients.

```{r}
opar <- par(no.readonly = T)
par(mfrow=c(1,2))
hist(res1,main = "Histogram")
qqnorm(res1)
qqline(res1,col ="blue")
shapiro.test(res1)
```
From the above result, we can conclude that the residuals appear to be white noise, while they are normally distributed. The shapiro test has a large p-value which does not indicate a rejection of normality assumption. Therefore, we can use Z-score for prediction intervals.

Now, prior to the prediction plot, we should notice that data we used to build the model were previously transformed with Box-cox transformation. To be consistent with original data, a de-transformation process is required. So we apply the inverse function of bc transformation to our predictions.

```{r}
#10 forcasts on transformed data and 95% CI
gaspred = predict(gas.fit1, n.ahead=10)


UpperB = gaspred$pred + 1.96*gaspred$se
max(UpperB)
LowerB = gaspred$pred - 1.96*gaspred$se
# transformation

ts.plot(ts(gas.bc),xlim = c(1,length(gas.bc)+20),ylim = c(min(gas.bc),max(UpperB)),main = "Forecast")

lines(233:242,(gaspred$pred + 1.96*gaspred$se), lty="dashed")

lines(233:242,(gaspred$pred - 1.96*gaspred$se),lty = "dashed")

points((length(gas.bc)+1):(length(gas.bc)+10),gaspred$pred,col="red")
```

```

```{r}
# 10 forecasts on original data and 95% CI
pred.org <- (exp(log(lambda*gaspred$pred+1)/lambda))
U = (exp(log(lambda*UpperB+1)/lambda))
L = (exp(log(lambda*LowerB+1)/lambda)) #transformation
```

```{r}
ts.plot(gas.train, xlim=c(1,length(gas.train)+10),ylim = c(1000,max(U)))

points((length(gas.train)+1):(length(gas.train)+10),pred.org,col="red")

lines(233:242,U,col = "Blue",lty = "dashed")

lines(233:242,L,col = "Blue", lty = "dashed")

```

```{r}
# Zoom the graph
ts.plot(gasorg,xlim=c(230,length(gas.train)+10),ylim = c(1500,max(U)))
lines(233:242,U,col="blue",lty = "dashed")
lines(233:242,L,col="blue",lty = "dashed")
points((length(gas.train)+1) : (length(gas.train)+10) , pred.org , col="red")
points((length(gas.train)+1) : (length(gas.train)+10), gas.test ,col = "green")

legend("topleft", pch = 1,col=c("red","green"), legend=c("Forecast value","True Value"))

lines(233:242,gas.test,col="gray",type="l")

```
```

Reference

Natural Gas Consumption. (2020, May 18). Retrieved from https://fred.stlouisfed.org/series/NATURALGASD11

Wikipedia contributors. (2020, June 4). Natural gas. In *Wikipedia, The Free Encyclopedia*. Retrieved 09:43, June 2, 2020, from https://en.wikipedia.org/w/index.php?title=Natural_gas&oldid=960709595