```
### STAT 5814 HW4/PROBLEM 3
### AUTHOR: SAYANTAN MAJUMDAR
### EMAIL: smxnv@mst.edu
### SNO: 12566087


library(TSA)
library(forecast)
library(snpar)
library(Rfit)

wd = "./Data/HW3_Data/"
setwd(wd)
ibm_stock = read.table('internet.txt')
internet = read.table('ibm.txt')
gasprices = read.table('gasprices.txt')

### IBM Stock Data Modeling
ibm_stock.fit = auto.arima(ibm_stock)
summary(ibm_stock.fit)
ibm_stock.transformed.fit <- auto.arima(ibm_stock,lambda="auto")
summary(ibm_stock.transformed.fit)

par(mfrow=c(2,2))
plot(ibm_stock.fit$residuals, ylab="Standardized Residuals",
type='l', main='Standardized Residual Plot')
abline(h = 0)
hist(ibm_stock.fit$residuals, main="Model Residual Histogram",
xlab="Residual")
qqnorm(ibm_stock.fit$residuals, main="QQ Plot for Residuals")
qqline(ibm_stock.fit$residuals, col="red")
acf(ibm_stock.fit$residuals, main="IBM Stock Residual ACF")

shapiro.test(ibm_stock.fit$residuals)
runs.test(ibm_stock.fit$residuals, exact=TRUE)

par(mfrow=c(2,2))
plot(ibm_stock.transformed.fit$residuals, ylab="Standardized
Residuals", type='l', main='Standardized Residual Plot After
Transformation')
abline(h = 0)
hist(ibm_stock.transformed.fit$residuals, main="Transformed Model
Residual Histogram", xlab="Residual")
```

```
qqnorm(ibm_stock.transformed.fit$residuals, main="QQ Plot for
Residuals After Transformation")
qqline(ibm_stock.transformed.fit$residuals, col="red")
acf(ibm_stock.transformed.fit$residuals, main="Transformed IBM Stock
Residual ACF")

shapiro.test(ibm_stock.transformed.fit$residuals)
runs.test(ibm_stock.transformed.fit$residuals, exact=TRUE)

### Internet Data Modeling
internet.fit = auto.arima(internet)
summary(internet.fit)
internet.transformed.fit <- auto.arima(internet,lambda="auto")
summary(internet.transformed.fit)

### Gasprices Data Modeling
gasprices.fit = auto.arima(gasprices)
summary(gasprices.fit)
gasprices.transformed.fit <- auto.arima(gasprices,lambda="auto")
summary(gasprices.transformed.fit)

par(mfrow=c(2,2))
plot(gasprices.fit$residuals, ylab="Standardized Residuals",
type='l', main='Standardized Residual Plot')
abline(h = 0)
hist(gasprices.fit$residuals, main="Model Residual Histogram",
xlab="Residual")
qqnorm(gasprices.fit$residuals, main="QQ Plot for Residuals")
qqline(gasprices.fit$residuals, col="red")
acf(gasprices.fit$residuals, main="Gas Prices Residual ACF")

shapiro.test(gasprices.fit$residuals)
runs.test(gasprices.fit$residuals, exact=TRUE)

par(mfrow=c(2,2))
plot(gasprices.transformed.fit$residuals, ylab="Standardized
Residuals", type='l', main='Standardized Residual Plot After
Transformation')
abline(h = 0)
hist(gasprices.transformed.fit$residuals, main="Transformed Model
Residual Histogram", xlab="Residual")
qqnorm(gasprices.transformed.fit$residuals, main="QQ Plot for
Residuals After Transformation")
```

```
qqline(gasprices.transformed.fit$residuals, col="red")
acf(gasprices.transformed.fit$residuals, main="Transformed Gas Prices
Residual ACF")

shapiro.test(gasprices.transformed.fit$residuals)
runs.test(gasprices.transformed.fit$residuals, exact=TRUE)
```

IBM Stock Data:

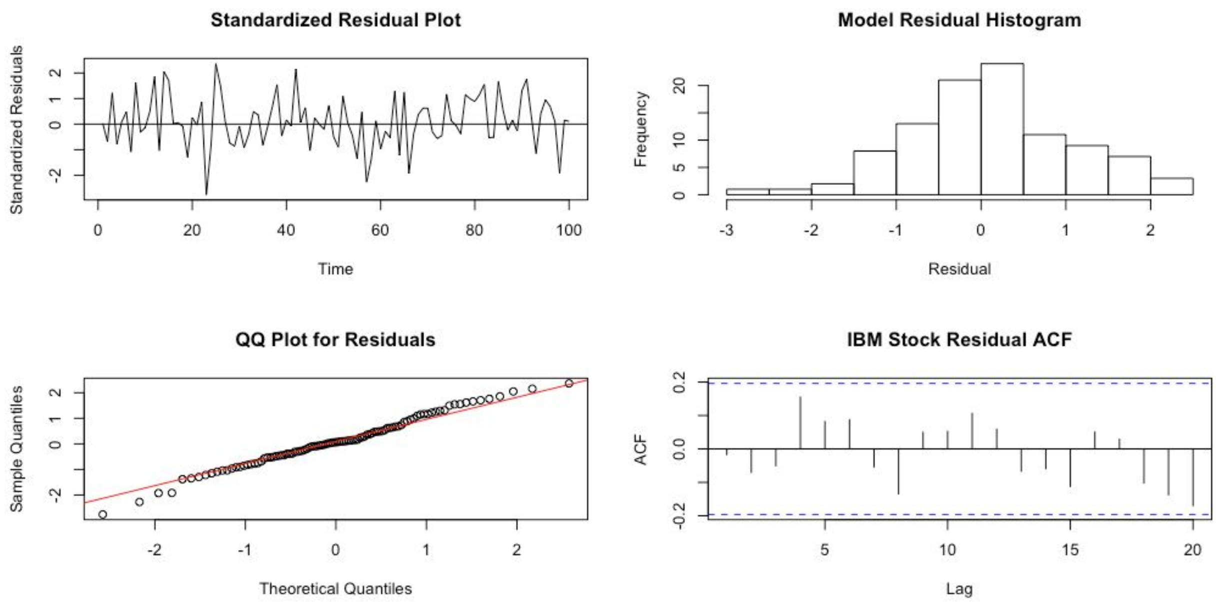
```
Series: ibm_stock
ARIMA(1,1,1)

Coefficients:
         ar1     ma1
      0.6504  0.5256
s.e.  0.0842  0.0896

sigma^2 estimated as 9.995:  log likelihood=-254.15
AIC=514.3   AICc=514.55   BIC=522.08

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE        ACF1
Training set 0.3035616 3.113754 2.405275 0.2805566 1.917463 0.5315228 -0.01715517
```

IBM Stock Transformed:

```
Series: ibm_stock
ARIMA(1,1,1)
Box Cox transformation: lambda= 0.3596253

Coefficients:
         ar1     ma1
      0.6486  0.4831
s.e.  0.0869  0.0978

sigma^2 estimated as 0.02025:  log likelihood=52.89
AIC=-99.79   AICc=-99.53   BIC=-92

Training set error measures:
                    ME     RMSE      MAE      MPE     MAPE      MASE       ACF1
Training set 0.2360619 3.107747 2.414814 0.228268 1.924939 0.5336308 0.01711577
```

Since the AIC and BIC are greatly reduced by applying Box Cox transformation, hence, the transformed model is better. We test this further by using residual analysis as follows:

IBM Stock Data:

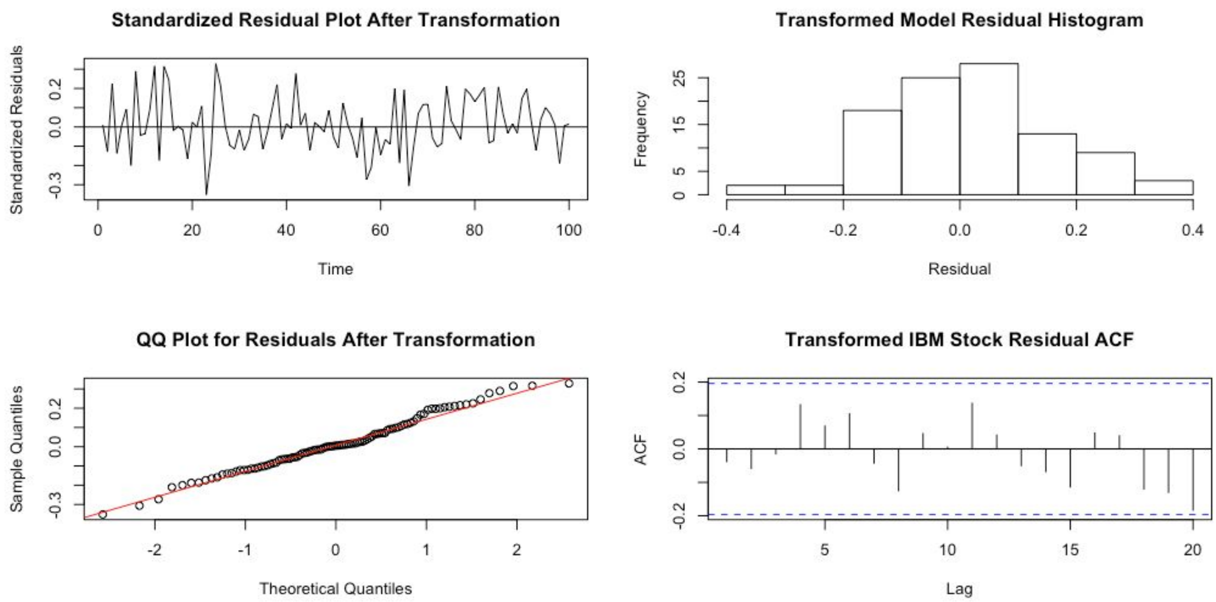


Shapiro-Wilk normality test
data: ibm_stock.residuals
W = 0.99057, p-value = 0.7107

Exact runs test
data: ibm_stock.fit$residuals
Runs = 54, p-value = 0.4813
alternative hypothesis: two.sided

Both the tests show that the model residuals are normally distributed. This is also observable from the above plots.

IBM Stock Transformed:



Shapiro-Wilk normality test
data: ibm_stock.transformed.fit$residuals
W = 0.98685, p-value = 0.4276

Exact runs test
data: ibm_stock.transformed.fit$residuals
Runs = 53, p-value = 0.6162
alternative hypothesis: two.sided

Here, we see that both the original and transformed model fit residuals follow a normal distribution. Therefore, we conclude that the transformed model will work better.

Internet Data:

Series: internet
ARIMA(1,1,0)

Coefficients:

```
         ar1
     -0.2035
s.e.  0.0901
```

sigma^2 estimated as 36116: log likelihood=-792.79

AIC=1589.59 AICc=1589.69 BIC=1595.15

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-5.436843	188.451	143.8879	-3.673256	16.09453	1.018055	0.004525616

Internet Data Transformed:

Series: internet
ARIMA(1,1,0)
Box Cox transformation: lambda= 1.999924

Coefficients:

	ar1
	-0.1873
s.e.	0.0907

sigma^2 estimated as 2.967e+10: log likelihood=-1603.12
AIC=3210.23 AICc=3210.34 BIC=3215.79

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-8.504401	186.379	141.4218	-3.960259	15.86436	1.000606	-0.03492769

Since the AIC and BIC are increasing after transformation, so we should choose the original data. Also, no normality test for the residuals is necessary to compare the original and transformed models.

Gas Prices Data:

Series: gasprices
ARIMA(1,1,0) with drift

Coefficients:

	ar1	drift
	0.4634	0.0122
s.e.	0.0742	0.0074

sigma^2 estimated as 0.002335: log likelihood=232.85
AIC=-459.7 AICc=-459.52 BIC=-450.79

Training set error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1

Training set -0.0003390976 0.04782376 0.03786608 0.01107526 1.40198 0.8814606 0.04844874

Gas Prices Transformed Data:

Series: gasprices
ARIMA(1,1,0) with drift
Box Cox transformation: lambda= 0.812215

Coefficients:

	ar1	drift
	0.4561	0.0103
s.e.	0.0747	0.0061

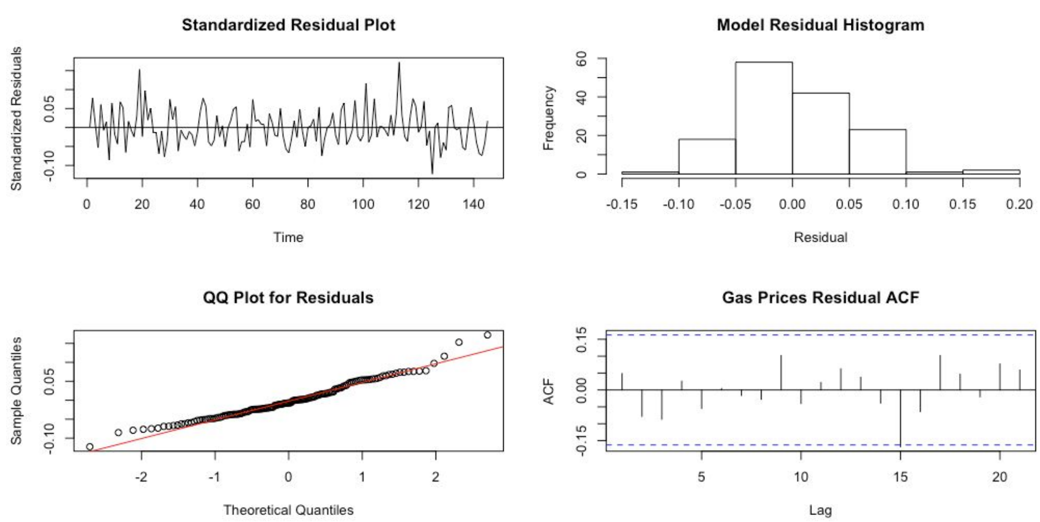
sigma^2 estimated as 0.001603: log likelihood=259.93
AIC=-513.86 AICc=-513.69 BIC=-504.95

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-0.000578663	0.04787401	0.0378926	0.003981294	1.40276	0.882078	0.05412511

Since the AIC and BIC are reducing for the transformed model, so it should be used instead of the original model. We test this further by using residual analysis as follows:

Gas Prices Data:

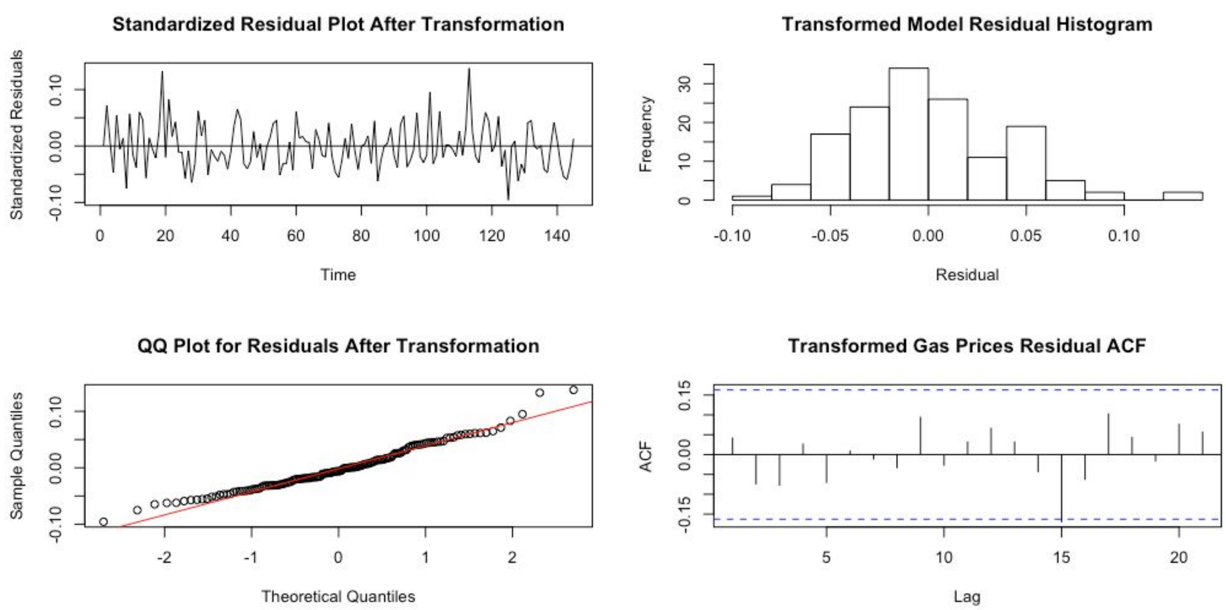


Shapiro-Wilk normality test
data: gasprices.fit$residuals
W = 0.97365, p-value = 0.006756

Exact runs test
data: gasprices.fit$residuals
Runs = 69, p-value = 0.5051
alternative hypothesis: two.sided

Since the p-value < 0.05 (for Shapiro-Wilk test), so the residuals are not normally distributed and there is systematic error present in the model fit. However, for the runs test, p-value > 0.05 which signifies that the residuals are normally distributed. This is also visible in the residual plots.

Gas Prices Transformed Data:



Shapiro-Wilk normality test
data: gasprices.transformed.fit$residuals
W = 0.97226, p-value = 0.00485

Exact runs test
data: gasprices.transformed.fit$residuals
Runs = 69, p-value = 0.5051
alternative hypothesis: two.sided

Since the p-value < 0.05 (for Shapiro-Wilk test), so the residuals are not normally distributed and there is systematic error present in the model fit. However, for the runs test, p-value > 0.05 which signifies that the residuals are normally distributed. This is also visible in the residual plots.