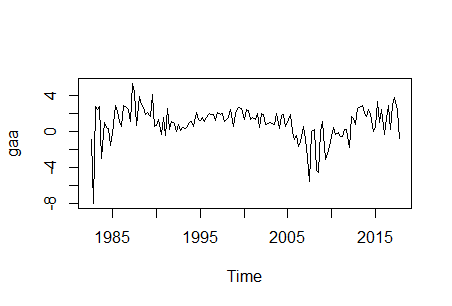
**Name: Anthony Ayebiahwe Predicting the Growth in Ann Arbor Housing Prices**

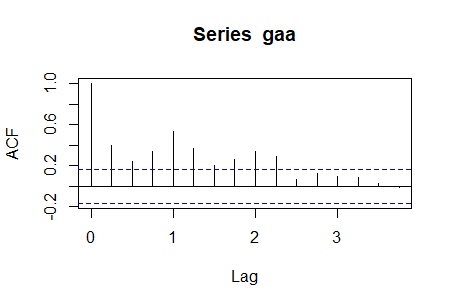
1)a) Plot of gaa over time is attached

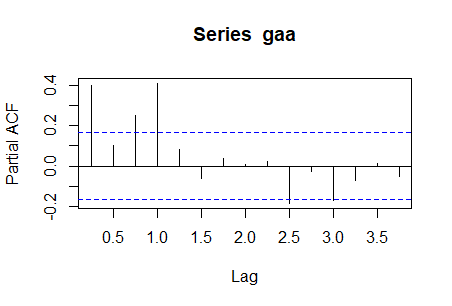


b)

|  |
| --- |
| Min. 1st Qu. Median Mean 3rd Qu. Max.  -8.0058 0.1698 1.0845 0.9050 1.9936 5.2983 |
| Standard deviation: 1.827251  Kurtosis: 7.621744  Skewness: -1.491794 |
| |  | | --- | |  | |

c) Plot of ACF and PACF are attached





d) Box-Ljung test

data: gaa

X-squared = 22.949, df = 1, p-value = 1.663e-06

> Box.test(gaa, lag = 2, type ="Ljung")

Box-Ljung test

data: gaa

X-squared = 31.641, df = 2, p-value = 1.347e-07

> Box.test(gaa, lag = 3, type ="Ljung")

Box-Ljung test

data: gaa

X-squared = 48.074, df = 3, p-value = 2.053e-10

> Box.test(gaa, lag = 4, type ="Ljung")

Box-Ljung test

data: gaa

X-squared = 89.946, df = 4, p-value < 2.2e-16

> Box.test(gaa, lag = 5, type ="Ljung")

Box-Ljung test

data: gaa

X-squared = 110.08, df = 5, p-value < 2.2e-16

> Box.test(gaa, lag = 6, type ="Ljung")

Box-Ljung test

data: gaa

X-squared = 115.98, df = 6, p-value < 2.2e-16

> Box.test(gaa, lag = 7, type ="Ljung")

Box-Ljung test

data: gaa

X-squared = 126.52, df = 7, p-value < 2.2e-16

> Box.test(gaa, lag = 8, type ="Ljung")

Box-Ljung test

data: gaa

X-squared = 143.58, df = 8, p-value < 2.2e-16

> Box.test(gaa, lag = 9, type ="Ljung")

Box-Ljung test

data: gaa

X-squared = 156.27, df = 9, p-value < 2.2e-16

> Box.test(gaa, lag = 10, type ="Ljung")

Box-Ljung test

data: gaa

X-squared = 156.92, df = 10, p-value < 2.2e-16

> Box.test(gaa, lag = 11, type ="Ljung")

Box-Ljung test

data: gaa

X-squared = 159.53, df = 11, p-value < 2.2e-16

> Box.test(gaa, lag = 11, type ="Ljung")

Box-Ljung test

data: gaa

X-squared = 159.53, df = 11, p-value < 2.2e-16

> Box.test(gaa, lag = 12, type ="Ljung")

Box-Ljung test

data: gaa

X-squared = 161.05, df = 12, p-value < 2.2e-16

> Box.test(gaa, lag = 13, type ="Ljung")

Box-Ljung test

data: gaa

X-squared = 162.31, df = 13, p-value < 2.2e-16

> Box.test(gaa, lag = 14, type ="Ljung")

Box-Ljung test

data: gaa

X-squared = 162.4, df = 14, p-value < 2.2e-16

> Box.test(gaa, lag = 15, type ="Ljung")

Box-Ljung test

data: gaa

X-squared = 162.42, df = 15, p-value < 2.2e-16

e) From the ACF, there are 9 positive spikes(9 lags) and after that smoothly decays to zero. From the PACF, the 1st and 4th have highest positive spikes. Moreover, from the PACF, it seems that there are limited number of spikes different from zero. So, I can entertain AR(2) or even as high as AR(5). Moreover, for the MA, I will take the 4 highest spikes in ACF and entertain MA(4) or even go as high as MA(9). I will choose Arma(2,4) or Arma(5,9) models. All the Q-stats have p-values that are less than 0.05, hence there is a time dependence in our data.

f) z test of coefficients:

Estimate Std. Error z value Pr(>|z|)

ar1 0.3167698 0.0813578 3.8935 9.879e-05 \*\*\*

ar2 0.0023433 0.0979338 0.0239 0.980911

ar3 0.3084376 0.0940605 3.2791 0.001041 \*\*

intercept 0.8483788 0.3522536 2.4084 0.016021 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

g) z test of coefficients:

Estimate Std. Error z value Pr(>|z|)

ar1 0.190745 0.074531 2.5593 0.01049 \*

ar2 -0.030962 0.084794 -0.3651 0.71501

ar3 0.108692 0.088833 1.2236 0.22112

ar4 0.534865 0.085211 6.2770 3.452e-10 \*\*\*

intercept 0.825897 0.554594 1.4892 0.13644

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

h) z test of coefficients:

Estimate Std. Error z value Pr(>|z|)

ma1 0.256576 0.084353 3.0417 0.002353 \*\*

ma2 0.116185 0.127996 0.9077 0.364024

ma3 0.201742 0.157552 1.2805 0.200377

ma4 0.461815 0.112687 4.0982 4.164e-05 \*\*\*

ma5 0.129603 0.147956 0.8760 0.381053

intercept 0.861600 0.272582 3.1609 0.001573 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

I) z test of coefficients:

Estimate Std. Error z value Pr(>|z|)

ar1 0.41604 0.16742 2.4849 0.01296 \*

ar2 -0.31756 0.17042 -1.8634 0.06241 .

ar3 0.19866 0.11367 1.7478 0.08051 .

ar4 0.48618 0.10712 4.5387 5.659e-06 \*\*\*

ma1 -0.25057 0.17920 -1.3983 0.16203

ma2 0.29290 0.16353 1.7911 0.07327 .

intercept 0.81800 0.52532 1.5571 0.11944

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

J)

Box-Ljung test

data: gaa.ar3.fit$residuals

X-squared = 22.519, df = 1, **p-value = 2.081e-06**

> Box.test(gaa.ar3.fit$residuals, lag = 5, type = c("Ljung"), fitdf = 3)

Box-Ljung test

data: gaa.ar3.fit$residuals

X-squared = 24.34, df = 2, **p-value = 5.185e-06**

> Box.test(gaa.ar3.fit$residuals, lag = 6, type = c("Ljung"), fitdf = 3)

Box-Ljung test

data: gaa.ar3.fit$residuals

X-squared = 25.748, df = 3, **p-value = 1.077e-05**

> Box.test(gaa.ar3.fit$residuals, lag = 7, type = c("Ljung"), fitdf = 3)

Box-Ljung test

data: gaa.ar3.fit$residuals

X-squared = 25.949, df = 4, p**-value = 3.24e-05**

> Box.test(gaa.ar3.fit$residuals, lag = 8, type = c("Ljung"), fitdf = 3)

Box-Ljung test

data: gaa.ar3.fit$residuals

X-squared = 28.182, df = 5**, p-value = 3.354e-05**

> Box.test(gaa.ar3.fit$residuals, lag = 9, type = c("Ljung"), fitdf = 3)

Box-Ljung test

data: gaa.ar3.fit$residuals

X-squared = 34.972, df = 6, **p-value = 4.364e-06**

> Box.test(gaa.ar3.fit$residuals, lag = 10, type =c("Ljung"), fitdf = 3)

Box-Ljung test

data: gaa.ar3.fit$residuals

X-squared = 38.386, df = 7, **p-value = 2.559e-06**

> Box.test(gaa.ar3.fit$residuals, lag = 11, type = c("Ljung"), fitdf = 3)

Box-Ljung test

data: gaa.ar3.fit$residuals

X-squared = 38.643, df = 8, **p-value = 5.725e-06**

> Box.test(gaa.ar3.fit$residuals, lag = 12, type = c("Ljung"), fitdf = 3)

Box-Ljung test

data: gaa.ar3.fit$residuals

X-squared = 38.809, df = 9, **p-value = 1.247e-05**

> Box.test(gaa.ar3.fit$residuals, lag = 13, type = c("Ljung"), fitdf = 3)

Box-Ljung test

data: gaa.ar3.fit$residuals

X-squared = 39.526, df = 10, **p-value = 2.053e-05**

# The residuals have some time dependencies in them since not all spikes in the ACF and PACF are within the dashed lines. All the Q- stats have p-values lower than 5%, so we reject the null hypothesis of white noise.

k) Box-Ljung test

data: gaa.ar4.fit$residuals

X-squared = 4.1013, df = 1, **p-value = 0.04285**

> Box.test(gaa.ar4.fit$residuals, lag = 6, type = c("Ljung"), fitdf = 4)

Box-Ljung test

data: gaa.ar4.fit$residuals

X-squared = 4.3376, df = 2, **p-value = 0.1143**

> Box.test(gaa.ar4.fit$residuals, lag = 7, type = c("Ljung"), fitdf = 4)

Box-Ljung test

data: gaa.ar4.fit$residuals

X-squared = 4.9769, df = 3, **p-value = 0.1735**

> Box.test(gaa.ar4.fit$residuals, lag = 8, type = c("Ljung"), fitdf = 4)

Box-Ljung test

data: gaa.ar4.fit$residuals

X-squared = 5.4049, df = 4, **p-value = 0.2482**

> Box.test(gaa.ar4.fit$residuals, lag = 9, type = c("Ljung"), fitdf = 4)

Box-Ljung test

data: gaa.ar4.fit$residuals

X-squared = 8.7324, df = 5, **p-value = 0.1202**

> Box.test(gaa.ar4.fit$residuals, lag = 10, type = c("Ljung"), fitdf = 4)

Box-Ljung test

data: gaa.ar4.fit$residuals

X-squared = 10.478, df = 6, **p-value = 0.1059**

> Box.test(gaa.ar4.fit$residuals, lag = 11, type =c("Ljung"), fitdf = 4)

Box-Ljung test

data: gaa.ar4.fit$residuals

X-squared = 11.227, df = 7, **p-value = 0.129**

> Box.test(gaa.ar4.fit$residuals, lag = 12, type = c("Ljung"), fitdf = 4)

Box-Ljung test

data: gaa.ar4.fit$residuals

X-squared = 13.854, df = 8, **p-value = 0.08564**

> Box.test(gaa.ar4.fit$residuals, lag = 13, type = c("Ljung"), fitdf = 4)

Box-Ljung test

data: gaa.ar4.fit$residuals

X-squared = 14.275, df = 9, **p-value = 0.1129**

# Only lag 5 has p-value less than 0.05 and statistical significant. The residuals appear to be white noise since the ACF and PACF spikes are inside the blue line.

L) The residuals seem like it has some time dependence, since the ACF and PACF has some spikes that are outside the blue line.

Box-Ljung test

data: gaa.ma5.fit$residuals

X-squared = 3.5256, df = 1, **p-value = 0.06043**

> Box.test(gaa.ma5.fit$residuals, lag = 7, type = c("Ljung"), fitdf = 5)

Box-Ljung test

data: gaa.ma5.fit$residuals

X-squared = 6.3683, df = 2, p-value = **0.04141**

> Box.test(gaa.ma5.fit$residuals, lag = 8, type = c("Ljung"), fitdf = 5)

Box-Ljung test

data: gaa.ma5.fit$residuals

X-squared = 13.715, df = 3, p-value = **0.00332**

> Box.test(gaa.ma5.fit$residuals, lag = 9, type = c("Ljung"), fitdf = 5)

Box-Ljung test

data: gaa.ma5.fit$residuals

X-squared = 18.37, df = 4, p-value = **0.001044**

> Box.test(gaa.ma5.fit$residuals, lag = 10, type = c("Ljung"), fitdf =5)

Box-Ljung test

data: gaa.ma5.fit$residuals

X-squared = 20.724, df = 5, p-value = **0.0009132**

> Box.test(gaa.ma5.fit$residuals, lag = 11, type = c("Ljung"), fitdf =5)

Box-Ljung test

data: gaa.ma5.fit$residuals

X-squared = 20.939, df = 6, p-value = **0.001882**

> Box.test(gaa.ma5.fit$residuals, lag = 12, type =c("Ljung"), fitdf = 5)

Box-Ljung test

data: gaa.ma5.fit$residuals

X-squared = 21.506, df = 7, p-value = **0.003089**

> Box.test(gaa.ma5.fit$residuals, lag = 13, type = c("Ljung"), fitdf =5)

Box-Ljung test

data: gaa.ma5.fit$residuals

X-squared = 21.565, df = 8, p-value = **0.005789**

# 6 of the lags highlighted in red have p-value less than 0.05, hence they are significant.

M) The residuals appear to be white noise since the ACF and PACF spikes are inside the blue line. Also, two of the lags highlighted in red have p-values less than 0.05, hence they are significant.

Box-Ljung test

data: gaa.arma42$residuals

X-squared = 5.4391, df = 1, **p-value = 0.01969**

> Box.test(gaa.arma42$residuals, lag = 8, type = c("Ljung"), fitdf = 6)

Box-Ljung test

data: gaa.arma42$residuals

X-squared = 5.656, df = 2, **p-value = 0.05913**

> Box.test(gaa.arma42$residuals, lag = 9, type = c("Ljung"), fitdf = 6)

Box-Ljung test

data: gaa.arma42$residuals

X-squared = 7.403, df = 3, **p-value = 0.0601**

> Box.test(gaa.arma42$residuals, lag = 10, type = c("Ljung"), fitdf = 6)

Box-Ljung test

data: gaa.arma42$residuals

X-squared = 8.2935, df = 4, p-value = **0.0814**

> Box.test(gaa.arma42$residuals, lag = 11, type = c("Ljung"), fitdf =6)

Box-Ljung test

data: gaa.arma42$residuals

X-squared = 9.1834, df = 5, p-value = **0.102**

> Box.test(gaa.arma42$residuals, lag = 12, type = c("Ljung"), fitdf =6)

Box-Ljung test

data: gaa.arma42$residuals

X-squared = 12.599, df = 6, p-value = **0.04987**

> Box.test(gaa.arma42$residuals, lag = 13, type =c("Ljung"), fitdf = 6)

Box-Ljung test

data: gaa.arma42$residuals

X-squared = 13.474, df = 7, p-value = **0.06137**

N)

gaa.rv.ar3

[1] 2.542388

> gaa.rv.ar4

[1] 1.987478

> gaa.rv.ma5

[1] 2.270238

> gaa.rv.arma42

[1] 1.938439

0) AIC(gaa.ar3.fit)

[1] 545.973

> BIC(gaa.ar3.fit)

[1] 560.7521

> AIC(gaa.ma5.fit)

[1] 534.4004

> BIC(gaa.ma5.fit)

[1] 555.0912

> AIC(gaa.ar4.fit)

[1] 514.2899

> BIC(gaa.ar4.fit)

[1] 532.0249

> AIC(gaa.arma42)

[1] 515.086

> BIC(gaa.arma42)

[1] 538.7326

P) The AIC and BIC of AR(4) is very low compared to the rest of the of the models; even though the residual variance is higher than that of the ARMA(4,2), hence I will select the AR(4) model.

q) Estimated model for AR(4)

AR(4)= 0.1624215+ 0.190745Yt−1 -0.030962Yt−2+ 0.108692Yt−3+ 0.534865Yt−4

σ2ε= 1.409779^2=**1.987477**

Forecast method: ARIMA(4,0,0) with non-zero mean

Model Information:

Call:

arima(x = gaa, order = c(4, 0, 0))

Coefficients:

ar1 ar2 ar3 ar4 intercept

0.1907 -0.0310 0.1087 0.5349 0.8259

s.e. 0.0745 0.0848 0.0888 0.0852 0.5546

sigma^2 estimated as 1.987: log likelihood = -251.14, aic = 514.29

Error measures:

ME RMSE MAE MPE MAPE MASE ACF1

Training set 0.05077859 1.409779 1.036463 -100.8137 419.8196 0.8467137 -0.03793648

Forecasts:

Point Forecast Lo 80 Hi 80 Lo 95 Hi 95

2018 Q1 1.829775 0.02307001 3.63648 -0.9333418 4.592892

r) f∗t,1= 0.1624215+ 1.829775Yt

Forecast uncertainties= **1.987477**

Actual forecast is

F1= c + phi1\*Yt + phi2\*Yt\_1 + phi3\*Yt\_2 + phi4\*Yt\_3

F1=0.1624215+0.1907445\*-0.86290321-0.03096169\*2.31950601+0.1086917\*3.72147871

F1=0.1624215+-0.164594-0.0718+0.404494

F1=**0.3305215**

The 95% confidence interval of the 1 step ahead forecast is **-0.9333418 4.592892**

**Code:**

library(tseries)

library(sandwich)

library(car)

library(forecast)

library(moments)

library(lmtest)

# Get data and convert to time series

AA= data$aa

AA=ts(AA, start = c(1982,2), end = c(2017,4), frequency = 4)

# lOG of A, find the difference, and plot the growth rate.

lAA=log(AA)

gaa=diff(lAA,lag = 1)\*100

ts.plot(gaa)

# summary stats of gaa

sumry= summary(gaa)

sumry

sdv=sd(gaa)

sdv

kurt.gaa=kurtosis(gaa)

kurt.gaa

skew.gaa= skewness(gaa)

skew.gaa

#acf and pacf for gaa over 15 lags

acf(gaa, lag.max = 15)

acf(gaa, lag.max = 15, type = "partial")

# Q-stats for the growth rate

Box.test(gaa, lag = 1, type ="Ljung")

Box.test(gaa, lag = 2, type ="Ljung")

Box.test(gaa, lag = 3, type ="Ljung")

Box.test(gaa, lag = 4, type ="Ljung")

Box.test(gaa, lag = 5, type ="Ljung")

Box.test(gaa, lag = 6, type ="Ljung")

Box.test(gaa, lag = 7, type ="Ljung")

Box.test(gaa, lag = 8, type ="Ljung")

Box.test(gaa, lag = 9, type ="Ljung")

Box.test(gaa, lag = 10, type ="Ljung")

Box.test(gaa, lag = 11, type ="Ljung")

Box.test(gaa, lag = 11, type ="Ljung")

Box.test(gaa, lag = 12, type ="Ljung")

Box.test(gaa, lag = 13, type ="Ljung")

Box.test(gaa, lag = 14, type ="Ljung")

Box.test(gaa, lag = 15, type ="Ljung")

# Ar(3) process

gaa.ar3.fit=arima(gaa, order = c(3,0,0))

coeftest(gaa.ar3.fit)

#Ar(4) Process

gaa.ar4.fit=arima(gaa, order = c(4,0,0))

coeftest(gaa.ar4.fit)

#Estimate the MA(5) Process

gaa.ma5.fit = arima(gaa,order = c(0,0,5))

coeftest(gaa.ma5.fit)

# Estimate the ARMA(4,2)

gaa.arma42 = arima(gaa,order = c(4,0,2))

coeftest(gaa.arma42)

# acf and pacf up to 13 lags of residuals of AR(3)

# Getting the Q-stats from 4 through 13

acf(gaa.ar3.fit$residuals,lag.max = 13)

acf(gaa.ar3.fit$residuals,lag.max = 13, type = "partial")

Box.test(gaa.ar3.fit$residuals, lag = 4, type = c("Ljung"), fitdf = 3)

Box.test(gaa.ar3.fit$residuals, lag = 5, type = c("Ljung"), fitdf = 3)

Box.test(gaa.ar3.fit$residuals, lag = 6, type = c("Ljung"), fitdf = 3)

Box.test(gaa.ar3.fit$residuals, lag = 7, type = c("Ljung"), fitdf = 3)

Box.test(gaa.ar3.fit$residuals, lag = 8, type = c("Ljung"), fitdf = 3)

Box.test(gaa.ar3.fit$residuals, lag = 9, type = c("Ljung"), fitdf = 3)

Box.test(gaa.ar3.fit$residuals, lag = 10, type =c("Ljung"), fitdf = 3)

Box.test(gaa.ar3.fit$residuals, lag = 11, type = c("Ljung"), fitdf = 3)

Box.test(gaa.ar3.fit$residuals, lag = 12, type = c("Ljung"), fitdf = 3)

Box.test(gaa.ar3.fit$residuals, lag = 13, type = c("Ljung"), fitdf = 3)

#acf and pacf up to 13 lags of residuals of AR(4)

# Getting the Q-stats from 5 through 13

acf(gaa.ar4.fit$residuals,lag.max = 13)

acf(gaa.ar4.fit$residuals,lag.max = 13, type = "partial")

Box.test(gaa.ar4.fit$residuals, lag = 5, type = c("Ljung"), fitdf = 4)

Box.test(gaa.ar4.fit$residuals, lag = 6, type = c("Ljung"), fitdf = 4)

Box.test(gaa.ar4.fit$residuals, lag = 7, type = c("Ljung"), fitdf = 4)

Box.test(gaa.ar4.fit$residuals, lag = 8, type = c("Ljung"), fitdf = 4)

Box.test(gaa.ar4.fit$residuals, lag = 9, type = c("Ljung"), fitdf = 4)

Box.test(gaa.ar4.fit$residuals, lag = 10, type = c("Ljung"), fitdf = 4)

Box.test(gaa.ar4.fit$residuals, lag = 11, type =c("Ljung"), fitdf = 4)

Box.test(gaa.ar4.fit$residuals, lag = 12, type = c("Ljung"), fitdf = 4)

Box.test(gaa.ar4.fit$residuals, lag = 13, type = c("Ljung"), fitdf = 4)

#acf and pacf up to 13 lags of residuals of MA(5)

# Getting the Q-stats from 6 through 13

acf(gaa.ma5.fit$residuals,lag.max = 13)

acf(gaa.ma5.fit$residuals,lag.max = 13, type = "partial")

Box.test(gaa.ma5.fit$residuals, lag = 6, type = c("Ljung"), fitdf = 5)

Box.test(gaa.ma5.fit$residuals, lag = 7, type = c("Ljung"), fitdf = 5)

Box.test(gaa.ma5.fit$residuals, lag = 8, type = c("Ljung"), fitdf = 5)

Box.test(gaa.ma5.fit$residuals, lag = 9, type = c("Ljung"), fitdf = 5)

Box.test(gaa.ma5.fit$residuals, lag = 10, type = c("Ljung"), fitdf =5)

Box.test(gaa.ma5.fit$residuals, lag = 11, type = c("Ljung"), fitdf =5)

Box.test(gaa.ma5.fit$residuals, lag = 12, type =c("Ljung"), fitdf = 5)

Box.test(gaa.ma5.fit$residuals, lag = 13, type = c("Ljung"), fitdf =5)

#acf and pacf up to 13 lags of residuals of ARMA(4,2)

# Getting the Q-stats from 4 through 13

acf(gaa.arma42$residuals,lag.max = 13)

acf(gaa.arma42$residuals,lag.max = 13, type = "partial")

Box.test(gaa.arma42$residuals, lag = 7, type = c("Ljung"), fitdf = 6)

Box.test(gaa.arma42$residuals, lag = 8, type = c("Ljung"), fitdf = 6)

Box.test(gaa.arma42$residuals, lag = 9, type = c("Ljung"), fitdf = 6)

Box.test(gaa.arma42$residuals, lag = 10, type = c("Ljung"), fitdf = 6)

Box.test(gaa.arma42$residuals, lag = 11, type = c("Ljung"), fitdf =6)

Box.test(gaa.arma42$residuals, lag = 12, type = c("Ljung"), fitdf =6)

Box.test(gaa.arma42$residuals, lag = 13, type =c("Ljung"), fitdf = 6)

#Find the residual variance for each model

gaa.rv.ar3 = gaa.ar3.fit$sigma2

gaa.rv.ar3

gaa.rv.ar4 = gaa.ar4.fit$sigma2

gaa.rv.ar4

gaa.rv.ma5 = gaa.ma5.fit$sigma2

gaa.rv.ma5

gaa.rv.arma42 = gaa.arma42$sigma2

gaa.rv.arma42

#Calculate the AIC and BIC for each model

#AR(3)

AIC(gaa.ar3.fit)

BIC(gaa.ar3.fit)

#MA(5)

AIC(gaa.ma5.fit)

BIC(gaa.ma5.fit)

#AR(4)

AIC(gaa.ar4.fit)

BIC(gaa.ar4.fit)

#ARMA(4,2)

AIC(gaa.arma42)

BIC(gaa.arma42)