**Part 3**

**Code:**

##choose what tv show and name it OUR\_TITLE

OUR\_TITLE <-"The Wire"

##check if the show is on imdb

res.1<-search\_by\_title(OUR\_TITLE,type="series")

res.1

#Check if it has ratings

find\_by\_title(OUR\_TITLE, type="episode", season=1, episode=1)$imdbRating

#Run function store ratings in variable named x

Wire<-getTv(OUR\_TITLE,OUR\_YEAR = 2002)

x<-Wire$x

##fit a least squares model to the times series with season numbers as the covariate.

lm1 <- lm(x~factor(season)-1,data=Wire)

##report the summary of the least squares model

summary(lm1)

## plot the residuals of the fitted model

x <- lm1$residuals

ts.plot(x)

acf(x)

##find the best arma model for the residuals of the least squares model

max.order <- 10

AIC.matrix <- matrix(1000,nrow = max.order+1,ncol= max.order+1)

for(i in 1:(max.order+1)){

for(j in 1:(max.order+1)){

currentArima <- arima(x,order=c(i-1,0,j-1))

AIC.matrix[i,j] <- AIC(currentArima)

}

}

min.aic<- which(AIC.matrix == min(AIC.matrix), arr.ind=TRUE)

p=min.aic[,1]-1

q=min.aic[,2]-1

##create best model and plot the acf of the residuals

best.model <- arima(x,order=c(p,0,q))

acf(residuals(best.model))

install.packages("nlme")

library(nlme)

##fit a generalized least squares model using the order of the best model we just fitted

gls1 <- gls(x~factor(season)-1,

data=Wire,

correlation = corARMA(value = .6\*coef(best.model)[1:11],p=7,q=4))

##look at the summary of the generalized least square and least squares model and compare them.

summary(gls1)

summary(lm1)

##install stargazer which helps me create a table of the regression results to compare the two models.

install.packages("stargazer")

library(stargazer)

regression\_results<-stargazer(lm1,gls1, type="text",title = "Regression Results")

Report:

For this part of the project I started by fitting a least squares model using the lm function with ratings as the dependent variable and the season numbers as the covariate. From this fitted model I then looked at a time series plot and the acf and pacf plots of the residuals of the fitted least squares model. From these plots it was clear to see that there was autocorrelation specifically at the lag of 12, which makes sense since it was following the trends of each season (every season was 12 episodes long). Since there was still autocorrelation in the acf and pacf plots it would make sense to try to find a better model to take into account of the seasonal trend of the data. From the summary of the fitted least squares model I saw that the coefficients of every season was statistically significant. I then used the function from part 1 of finding the order of the best ARMA model by storing all ARMA models of p=0 to 10, and q=0 to 10, and finding the ARMA model with the lowest AIC. From this I got a ARMA(7,4). I then fitted a generalized least squares model with the parameters of the best ARMA model.

Finally, I examined my results of the fitted general least squares model and compared them to my fitted least squares model. As seen in the regression results table below the coefficients of parameter are fairly similar however the standard errors of the gls model are much smaller than the standard errors of the lm model. The parameters of the gls model are more accurate than the lm model because of the smaller standard errors, which take into account the autocorrelation due to the seasons. In both models the the parameters are significant which shows what I believed to be true from the beginning that depending on the season the ratings are going to vary. One other thing is that the AIC of the gls model is smaller than the lm model which is desirable for fitting a good model.

>ts.plot(x) ## this shows the ts plot of the residuals fo the lm1 model



>acf(x) ## this shows the acf plot of the residuals fo the lm1 model



>acf(residuals(best.model)) ## this shows the acf plot of the residuals fo the best arma model



>regression\_results<-stargazer(lm1,gls1, type="text",title "Regression Results")

Regression Results

===========================================================

Dependent variable:

---------------------------------------

x

OLS generalized

least squares

(1) (2)

-----------------------------------------------------------

factor(season)1 8.731\*\*\* 8.734\*\*\*

(0.095) (0.011)

factor(season)2 8.683\*\*\* 8.687\*\*\*

(0.099) (0.009)

factor(season)3 8.883\*\*\* 8.913\*\*\*

(0.099) (0.008)

factor(season)4 8.900\*\*\* 8.868\*\*\*

(0.095) (0.009)

factor(season)5 8.930\*\*\* 9.022\*\*\*

(0.108) (0.018)

-----------------------------------------------------------

Observations 60 60

R2 0.999

Adjusted R2 0.998

Log Likelihood -1.390

Akaike Inf. Crit. 48.568 36.779

Bayesian Inf. Crit. 70.904

Residual Std. Error 0.343 (df = 55) 0.3606166

F Statistic 7,949.329\*\*\* (df = 5; 55)

===========================================================

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

###summary of gls model

>summary(gls1)

Generalized least squares fit by REML

Model: x ~ factor(season) - 1

Data: Wire

AIC BIC logLik

36.77901 70.90367 -1.389505

Correlation Structure: ARMA(7,4)

Formula: ~1

Parameter estimate(s):

Phi1 Phi2 Phi3 Phi4 Phi5 Phi6 Phi7

2.2364381 -2.7978233 2.0541953 -1.1867904 0.4928726 -0.1778961 -0.1301082

Theta1 Theta2 Theta3 Theta4

-2.7011434 3.2836260 -2.0575683 0.4750858

Coefficients:

Value Std.Error t-value p-value

factor(season)1 8.733834 0.011108176 786.2527 0

factor(season)2 8.687483 0.008810822 986.0014 0

factor(season)3 8.913081 0.008481360 1050.9024 0

factor(season)4 8.868400 0.008639950 1026.4411 0

factor(season)5 9.021881 0.017556372 513.8807 0

Correlation:

fct()1 fct()2 fct()3 fct()4

factor(season)2 -0.710

factor(season)3 0.311 -0.713

factor(season)4 -0.131 0.302 -0.708

factor(season)5 0.064 -0.138 0.325 -0.719

Standardized residuals:

Min Q1 Med Q3 Max

-1.7576391 -0.6484277 -0.1896740 0.4929168 2.0287488

Residual standard error: 0.3606166

Degrees of freedom: 60 total; 55 residual