Final Project 427

Torrence Savage and Muu

12/4/17

Unemployment and GDP Growth

Table of Contents:

Introduction:

Good morning everybody, today we will be presenting on Unemployment Rate and Gross Domestic Product Growth Rate. Our goals for this project are to look at Unemployment Rate and Gross Domestic Product as time series, forecast both, and model Unemployment Rate based on GDP growth.

Goals:

Compare Unemployment Rate and GDP

Build a model that can forecast both

Model UR based on GDP

Resources:

www.bea.gov

Extracted GDP growth using chained 2009 dollars from 1947 to 2016.

https://www.bls.gov/data/

R Studio

JMP Pro 13

Process:

Data Collection and Importing

First we just googled statistics data sets to find something we could analyze. We used data sets from www.bea.gov and www.bls.gov/data to examine Gross Domestic Product growth rate and Unemployment Rate. We got an Excel file that we then imported into R to further analyze. We also used JMP to look at possible model selections.

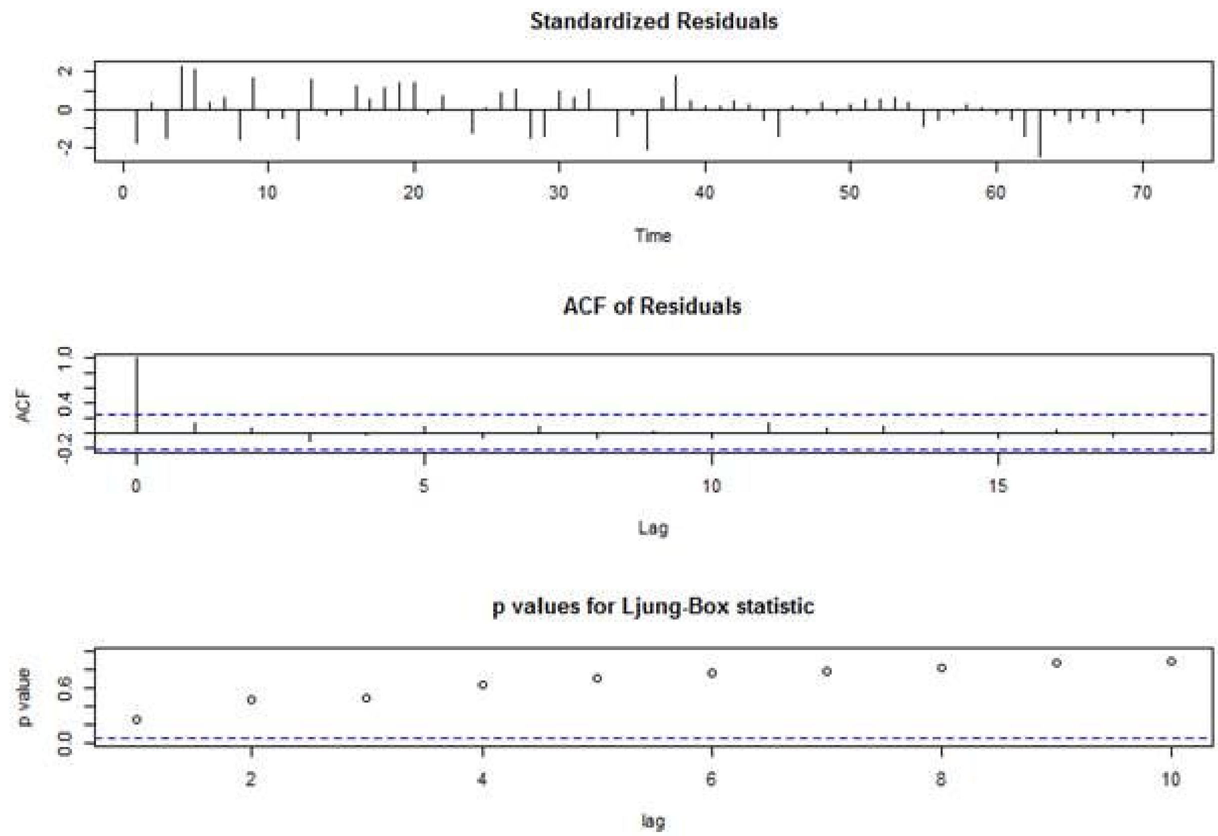
```
#Import Data
f <- (Final_Group_Project_Data)
f
final.data <- ts(f)
final.data

x <- c(final.data[,2])
x
y <- c(final.data[,3])
y
```

Time Series plots

Our time series plot of Growth of the Gross Domestic Product appears random and either stationary or with a slight negative trend, based on visual inspection with mean of approximately 3-4 %. Next, we examined it based on the ACF and PACF. They also indicated that the GDP Growth was stationary because the acf decayed rapidly, and therefore could possibly be modeled with an MA(q) process. We then used the R function auto.arima to look at a group of possible models. We found that the best model according to auto.arima was the white noise process of ARIMA(0,0,0), as we had originally thought. Finally we used the tsdiag function in R to look at the residuals to determine whether or not there was autocorrelation. The residual plot looked okay, the ACF had no significant values, and the Ljung Box statistic plot plot had all high p-values, indicating that there was no autocorrelation present. This was double checked in JMP with a model group comparison. JMP showed that the ARIMA(0,0,0) was process was the second best model, with the IMA(0,1,1) model being the best. This would make sense if there was a linear trend that needed to be differenced to truly get a stationary process.

Here are the results from our code along with JMP Output.

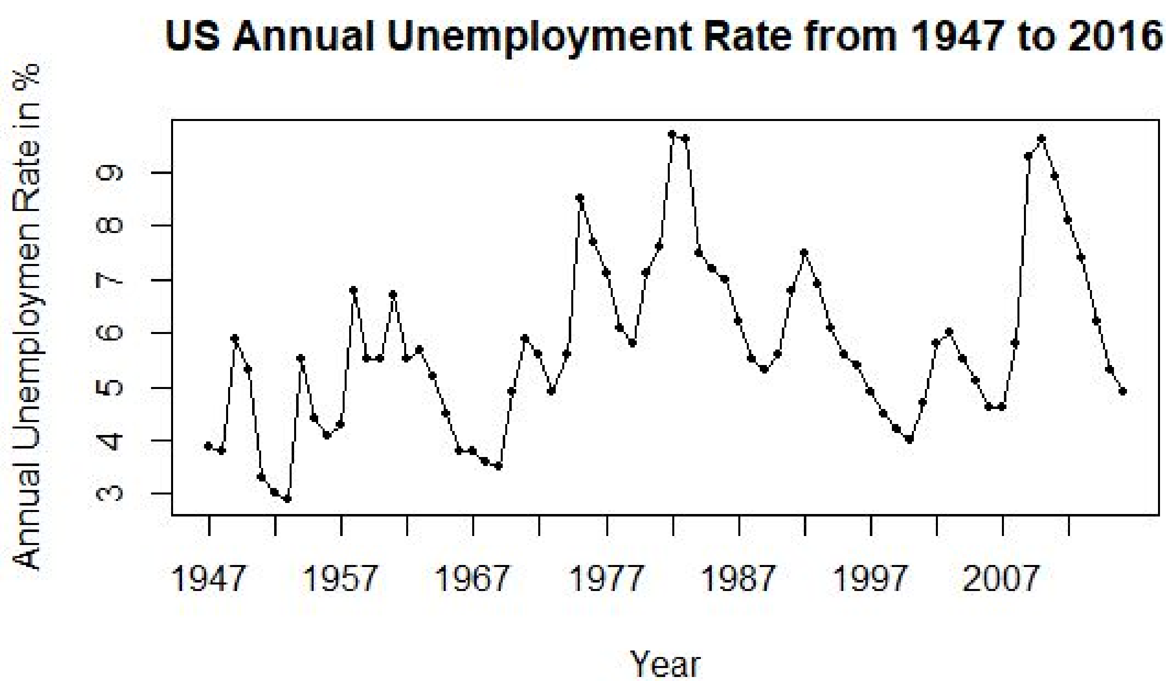
```
#GDP Plot
plot(final.data[1:70, 2],type = "o", pch = 16, cex = .5, xlab = 'Year',
   ylab = 'GDP Annual Rate in %', main = "US GDP Annual Rate from 1947 to 2016",
   xaxt = 'n')
axis(1, seq(1, 70, 5), final[seq(1, 70, 5), 1])

par(mfrow=c(1,2),oma=c(0,0,0,0))
acf(final.data[1:70, 2],lag.max = 25,type = "correlation",  main = "ACF for the Annual \nGDP
Rate")
pacf(final.data[1:70, 2])
```

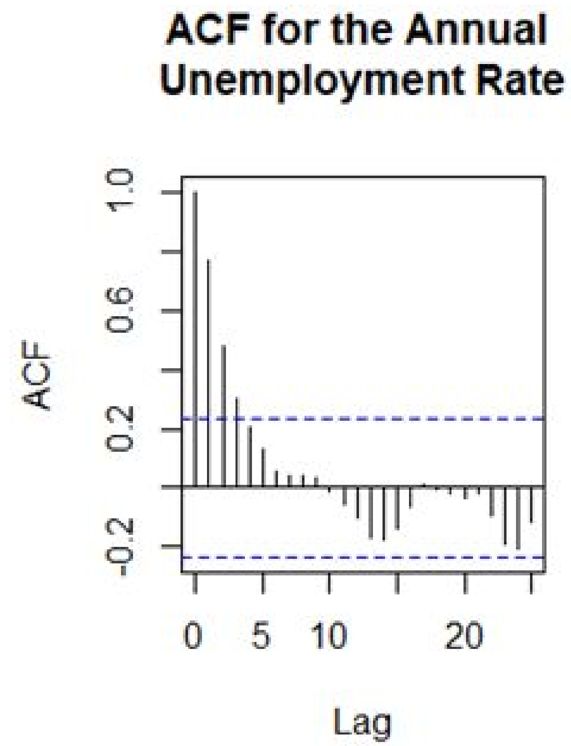


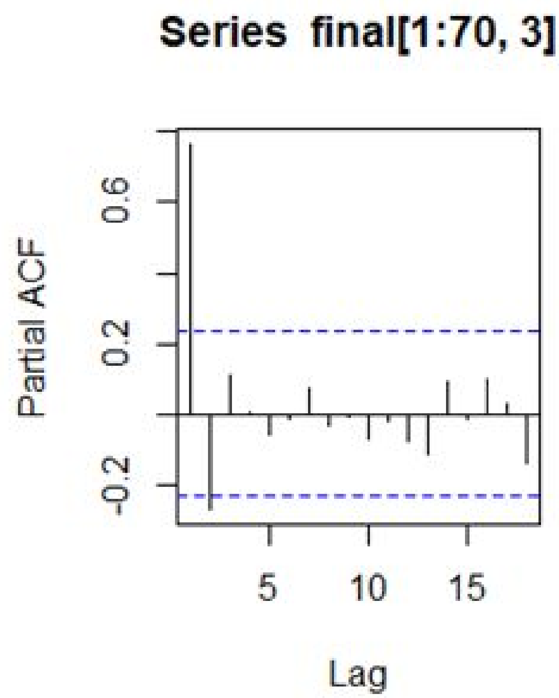
We did the same process for looking at Unemployment Rate, with different results. The time series plot looked to have a pattern of ups and downs but the the number of years between cycles is different, so it can’t be modeled seasonally. Also, it seems to be linearly trending upward, so a differencing order of one may be needed. The ACF has a gradual decay, and the PACF decays rapidly, so it should be modeled with an ARIMA process. We used the auto.arima process again to look at potentially good models and it gave us the IMA(0,1,2) process as the best model. We decided to double check with JMP again before we got too far, and this time JMP showed that as not even being a top 10 model. JMP showed that the ARIMA (2,1,1) model was the best.

```
#Unemployment Plot
plot(final.data[1:70, 3],type = "o", pch = 16, cex = .5, xlab = 'Year',
   ylab = 'Annual Unemploymen Rate in %', main = "US Annual Unemployment Rate from 1947
to 2016",
   xaxt = 'n')
axis(1, seq(1, 70, 3), final[seq(1, 70, 3), 1])
```

```
par(mfrow=c(1,2),oma=c(0,0,0,0))
acf(final[1:70, 3],lag.max = 25,type = "correlation",  main = "ACF for the Annual
\nUnemployment Rate")
pacf(final[1:70, 3])
```







GDP (moving average) based on ACF/PACF

Unemployment rate(ARIMA model)

Auto.arima easiest method of choosing model.

Forecasting:

Linear Model +AR model:

```
#Linear Model
linear_model<-lm(y~x)
linear_model
```

```
#AR Model Part
m<-linear_model$residuals
plot(m)
ar_part=ar(m)
ar_part$order
```

Conclusion:

GDP Growth rate is best modeled by a random white noise process.

There is a correlation between GDP growth and unemployment rate.

R.Code

```
#Import Data
f <- (Final_Group_Project_Data)
f
final.data <- ts(f)
final.data

x <- c(final.data[,2])
x
y <- c(final.data[,3])
y

#GDP Plot
plot(final.data[1:70, 2],type = "o", pch = 16, cex = .5, xlab = 'Year',
```

```
    ylab = 'GDP Annual Rate in %', main = "US GDP Annual Rate from 1947 to 2016",
    xaxt = 'n')
axis(1, seq(1, 70, 5), final[seq(1, 70, 5), 1])

par(mfrow=c(1,2),oma=c(0,0,0,0))
acf(final.data[1:70, 2],lag.max = 25,type = "correlation",  main = "ACF for the Annual \nGDP
Rate")
pacf(final.data[1:70, 2])

#c) Model Selection
#Automatic Selection Algorithm - Fast
auto.arima(final.data[1:60, 2], trace= TRUE, ic ="aicc", approximation = FALSE)

# Fit model to first few years of GDP data
gdp.fit.arima<- arima(final.data[1:60, 2],order =c(0,0,0))
gdp.fit.arima


#Fit the residuals
res.gdp.arima<-as.vector(residuals(gdp.fit.arima))
res.gdp.arima
#to obtain the fitted values we use the function fitted() from
#the forecast package
library(forecast)
fit.gdp.arima<-as.vector(fitted(gdp.fit.arima))
fit.gdp.arima

Box.test(res.gdp.arima, lag = 18, fitdf = 0, type = "Ljung")
#fitdf = p+q


#ACF and PACF of the Residuals
par(mfrow=c(1,2),oma=c(0,0,0,0))
acf(res.gdp.arima,lag.max=25,type="correlation",
  main="ACF of the Residuals \nof ARIMA(0, 0, 0) Model")
acf(res.gdp.arima, lag.max=25,type="partial",
  main="PACF of the Residuals \nof ARIMA(0, 0, 0) Model")
```

```
#d) Residual Analysis (4-in-1 plot of the residuals)
par(mfrow=c(2,2),oma=c(0,0,0,0))
qqnorm(res.gdp.arima,datax = TRUE,pch=16,xlab='Residual',main='')
qqline(res.gdp.arima,datax = TRUE)
plot(fit.gdp.arima, res.gdp.arima,pch=16,xlab='Fitted Value',ylab='Residual')
abline(h=0)
hist(res.gdp.arima,col="gray",xlab='Residual',main='')
plot(res.gdp.arima,type= "l",xlab='Observation Order',ylab='Residual')
points(res.gdp.arima,pch=16,cex=.5)
abline(h=0)




#Unemployment Plot
plot(final.data[1:70, 3],type = "o", pch = 16, cex = .5, xlab = 'Year',
   ylab = 'Annual Unemploymen Rate in %', main = "US Annual Unemployment Rate from 1947
to 2016",
   xaxt = 'n')
axis(1, seq(1, 70, 3), final[seq(1, 70, 3), 1])

par(mfrow=c(1,2),oma=c(0,0,0,0))
acf(final[1:70, 3],lag.max = 25,type = "correlation",  main = "ACF for the Annual
\nUnemployment Rate")
pacf(final[1:70, 3])


#c) Model Selection
#Automatic Selection Algorithm - Fast
auto.arima(final.data[1:60, 3], trace= TRUE, ic ="aicc", approximation = FALSE)
#Auto Algorithm - Slow but more accurate
auto.arima(final.data[1:60, 3],trace= TRUE, ic ="aicc", approximation = FALSE, stepwise =
FALSE)


# Fit (0,1,2) model to first few years of GDP data
```

```
unemploy1.fit.arima<- arima(final.data[1:60, 3],order =c(0,1,2))
unemploy1.fit.arima


#Fit the residuals(0,1,2)
res.unemploy1.arima<-as.vector(residuals(unemploy1.fit.arima))
res.unemploy1.arima
#to obtain the fitted values we use the function fitted() from
#the forecast package
library(forecast)
fit.unemploy1.arima<-as.vector(fitted(unemploy1.fit.arima))
fit.unemploy1.arima

Box.test(res.unemploy1.arima, lag = 18, fitdf = 0, type = "Ljung")
#fitdf = p+q


#ACF and PACF of the Residuals
par(mfrow=c(1,2),oma=c(0,0,0,0))
acf(res.unemploy1.arima,lag.max=25,type="correlation",
  main="ACF of the Residuals \nof ARIMA(0, 1, 2) Model")
acf(res.unemploy1.arima, lag.max=25,type="partial",
  main="PACF of the Residuals \nof ARIMA(0, 1, 2) Model")

#d) Residual Analysis (4-in-1 plot of the residuals)
par(mfrow=c(2,2),oma=c(0,0,0,0))
qqnorm(res.unemploy1.arima,datax = TRUE,pch=16,xlab='Residual',main='')
qqline(res.unemploy1.arima,datax = TRUE)
plot(fit.unemploy1.arima, res.unemploy1.arima,pch=16,xlab='Fitted Value',ylab='Residual')
abline(h=0)
hist(res.unemploy1.arima,col="gray",xlab='Residual',main='')
plot(res.unemploy1.arima,type= "l",xlab='Observation Order',ylab='Residual')
points(res.unemploy1.arima,pch=16,cex=.5)
abline(h=0)


# Fit (2,1,1)model to first few years of Unemployment data
```

```
unemploy2.fit.arima<- arima(final.data[1:60, 3],order =c(2,1,1))
unemploy2.fit.arima

#Fit the residuals(2,1,1)
res.unemploy2.arima<-as.vector(residuals(unemploy2.fit.arima))
res.unemploy2.arima
#to obtain the fitted values we use the function fitted() from
#the forecast package
library(forecast)
fit.unemploy2.arima<-as.vector(fitted(unemploy2.fit.arima))
fit.unemploy2.arima

Box.test(res.unemploy2.arima, lag = 18, fitdf = 0, type = "Ljung")
#fitdf = p+q


#ACF and PACF of the Residuals
par(mfrow=c(1,2),oma=c(0,0,0,0))
acf(res.unemploy2.arima,lag.max=25,type="correlation",
  main="ACF of the Residuals \nof ARIMA(2, 1, 1) Model")
acf(res.unemploy2.arima, lag.max=25,type="partial",
  main="PACF of the Residuals \nof ARIMA(2, 1, 1) Model")

#d) Residual Analysis (4-in-1 plot of the residuals)
par(mfrow=c(2,2),oma=c(0,0,0,0))
qqnorm(res.unemploy2.arima,datax = TRUE,pch=16,xlab='Residual',main='')
qqline(res.unemploy2.arima,datax = TRUE)
plot(fit.unemploy2.arima, res.unemploy2.arima,pch=16,xlab='Fitted Value',ylab='Residual')
abline(h=0)
hist(res.unemploy2.arima,col="gray",xlab='Residual',main='')
plot(res.unemploy2.arima,type= "l",xlab='Observation Order',ylab='Residual')
points(res.unemploy2.arima,pch=16,cex=.5)
abline(h=0)
```

ARIMA(2,0,2) with non-zero mean : 282.4774

```
ARIMA(0,0,0) with non-zero mean : 277.6548
ARIMA(1,0,0) with non-zero mean : 279.7928
ARIMA(0,0,1) with non-zero mean : 279.7887
ARIMA(0,0,0) with zero mean     : 343.6251
ARIMA(1,0,1) with non-zero mean : Inf

Best model: ARIMA(0,0,0) with non-zero mean

Series: final[1:60, 2]
ARIMA(0,0,0) with non-zero mean

Coefficients:
       mean
     3.4333
s.e.  0.3050

sigma^2 estimated as 5.676:  log likelihood=-136.72
AIC=277.44   AICc=277.65   BIC=281.63
```