Time Series Analysis of Unemployment Data

Arthur Hla, Quanming Huang, Colby Jordan, Jasmine Kellogg, Janavi Kumar
Prof. Sudeep Bapat

Fall 2018, UCSB

1 Abstract

In the labor market, understanding and predicting unemployment rates allows the working class to be prepared for any sudden job loss or market change. The objective of this project is to use our time series analysis methods in order to predict future unemployment rates by using past unemployment rates. We will difference and transform our data on unemployment in order to build a SARIMA (And explain more when we actually get the model). Then, we will perform diagnostic checks on our final model and forecast future data, through which we will obtain predicted values within a confidence level of 95% which will be near the true values of our original data set.

2 Introduction

With how crucial maintaining a steady job is for most of the population, understanding the changes of unemployment is extremely important so that workers can better prepare for any job loss. For many in the past, such as those living during the great depression, having a steady job means the difference between having food on the table and living on the streets.

The data in question concerns monthly unemployment rates for males aged 16 to 19 from 1948 to 1981 in the United States. We chose this time series because it has a sufficient number of data points (408) for analysis and can provide us with insight into unemployment over the years. Our data has two variables, time and the number of unemployed males (in thousands). For forecasting purposes, we remove the last year of data (the last 12 data points) so we may predict these values in 1981 using our chosen time series model. This brings our data points down from 408 to 396. The labour market is constantly changing and the data from this time series describes the labour market during period of 1948-1981. In addition, by focusing on men between the ages of 16 to 19, we can understand how the labor market is changing for a youth population which becomes more focused on careers at an earlier age every year. By looking at the data, we're able to get a feel for the state of the job market in the mid 1900's and compare our findings to the job market today.

Our procedure was as such: First, we plotted the original data and did an initial analysis of our findings. We noted a seasonal trend as well as a lack of stationarity. To correct this, we used a Box-Cox transformation and the differencing method. After doing a Box-Cox transformation, it was clear that there was still an upward trend and seasonality, so we continued to difference the data. First, we differenced at lag 12 as our seasonal component is 12 due to monthly data. This initial differencing addressed the seasonality aspect. However, the data was still not stationary. We differenced again, this time at lag 1. This gave us our transformed data which was stationary. After transforming the data, we used ACF and PACF analysis as well BIC and AIC analysis to fit our model. We chose two models to take a closer look at, both of which were causal and invertible. We then used Diagnostic techniques such as the Shapiro Test, Box Pierce Test, and Ljung Box Test to assess the accuracy and performance of the two models. Both models passed all three tests as well as QQ and histogram tests for normality. Thus, we proceeded to pick our final model. Out of our two models, we chose the one with fewer coefficients which was SARIMA(0,1,1) x $(1,1,0)_{12}$. This model most accurately allowed us to forecast future unemployment values.

3 Initial Time Series

3.1 Initial Data Plot

First, we graph the initial data of unemployment rates.

Looking at our plot below, we see several issues keeping the series from achieving stationarity. First, there is a general upward trend and increase in variance over the entire given time period. This is expected; we can interpret this as the steady growth of population in the U.S. over the years, and in turn an increase in number of potential unemployed 16-19 year old males. Further, there is a seasonal spike every year in June, giving a peak number of unemployment annually. At first glance, this may seem confusing, as most student aged males participate in summer jobs that they may have not hold during the school year. However, we must take into consideration how unemployment is defined; an unemployed person is someone actively looking for a job, not just a person who does not have a job. When an individual has not been looking for a job for over 4 weeks, they are taken out of the labor force and are no longer considered unemployed. Therefore, during the school year, many males aged 16 to 19 will be taken out of the unemployment count due to inactivity; however, in June, many students will start to look for jobs, and will be counted again as unemployed individuals, which is what the June spike indicates. This unemployment count then immediately declines as these young men find and begin summer jobs. In order to accurately represent the true unemployment, we will need to remove this seasonality and the general upward trend.

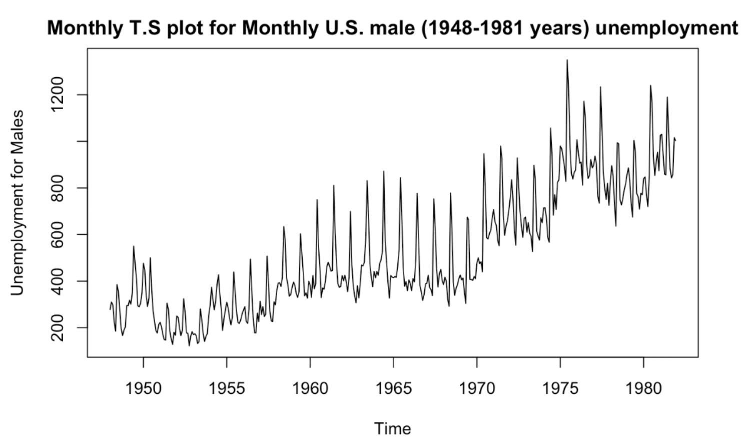


Figure 1: Graph of Initial Raw Data

[1]

[1]The y variable our initial data, as shown in Figure 1, is in the unit of thousands of men. Thus, our data concerns between 200,000 and 1.2 million men. This is the same for our Seasonal plot in Figure 3 and the Data portion of our Decomposition Plot in Figure 2

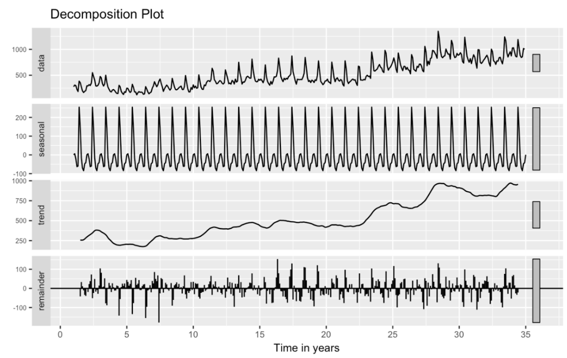


Figure 2: Decomposition Plot of Initial Data

3.2 Decomposition Plot

Observing our decomposition plot above, we confirm our suspicions listed above. We see in the seasonality section that there is a consistent spike every year. This, as explained above, is due to the increase of teenagers searching for jobs in the beginning summer months explained above. In addition, there is a clear upward trend over the entire time period due to population increase. Both of these aspects show how our model is not stationary. Thus, we will need to transform our data in order to remove both our trend and the seasonal component of the data.

3.3 Seasonal Model

By looking at the seasonal plot below, we see that the unemployment numbers hover around 1000 units for most of the year, and spike in early summer.

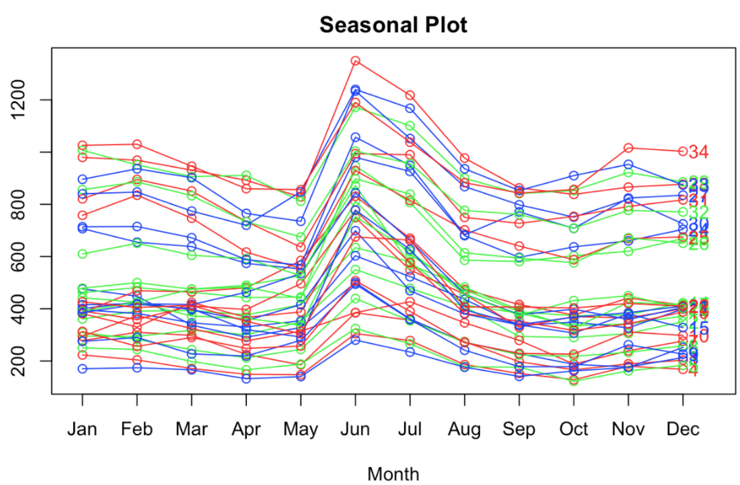


Figure 3: Seasonal Plot of Initial Data

4 Time Series Transformation

In order to create stationary data out of our original time series, we must stabilize our variance using the Box-Cox transformation. From there, we can remove the seasonality and trend of our series.

4.1 Box-Cox Transformation

In order to receive our λ for our transformation, we will do a Box-Cox Transformation. Figures 4 and 5 shows our plot for our Box-Cox transformation as well as a graph of our data after undergoing the transformation. As we can see in Figure 4, 0 does not lie in our interval which means our transformation indeed works. In addition, we have calculated that our λ used for our Box-Cox transformation is .2626263. This means we have our new transformed data as $V_t = X_t^{.2626263}$. We can see this transformed data in Figure 5.

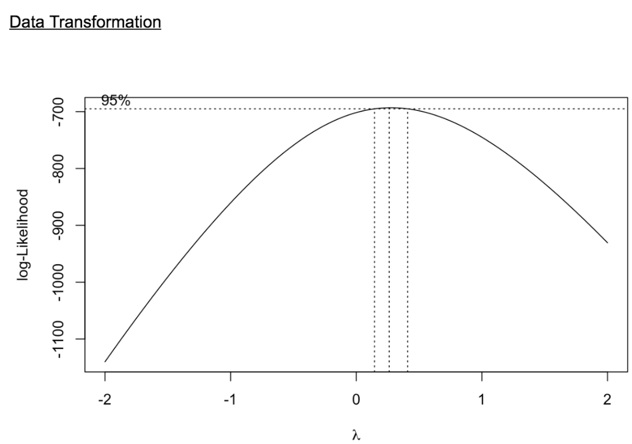


Figure 4: Graph of Box-Cox Plot

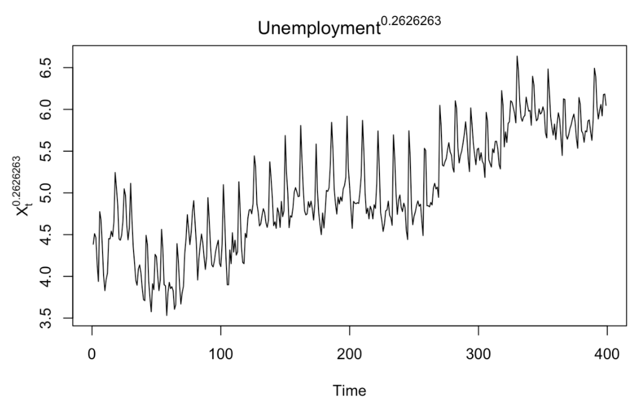


Figure 5: Graphs of Data Post Box-Cox Transformation and corresponding ACF

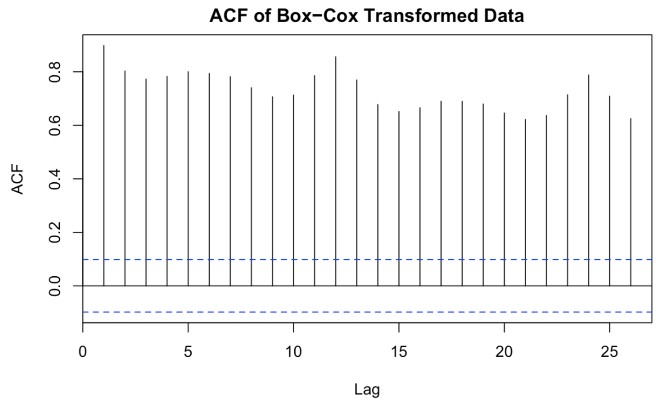


Figure 6: ACF of Box-Cox Transformed Data

However, in Figure 6 we plot the ACF of our Box-Cox transformed data. As we can see, our ACF does not lag or cut off. Therefore, we need to do further steps. In order to make our data stationary, our next step is differencing our data.

4.2 Differencing and Detrending

To remove the seasonality of our trend, we will difference our model. As we can see a spike in unemployment each June, it follows that our data is seasonal for a period of every 12 months. Thus, we use the formula $W_t = \nabla_d V_t = V_t - V_{t-d}$ with $d = 12$. In Figure 8, we plot our differenced data.

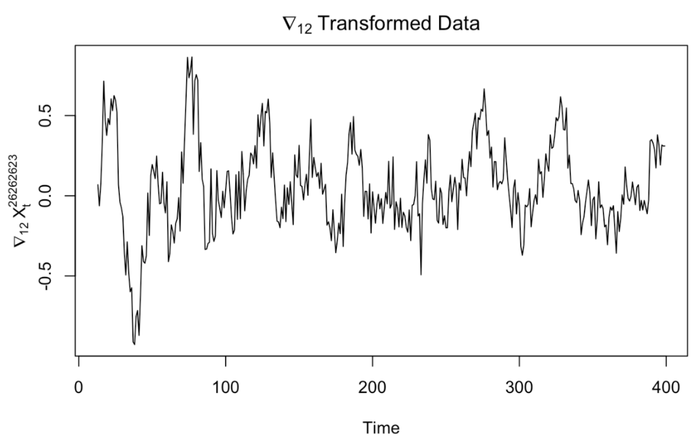


Figure 7: Graph of Deseasonalized Data Post Differencing at Lag = 12

After differencing at lag 12, we see that the data is no longer affected by seasonality, but still shows a slight upward trend. In order to get rid of this trend, we differenced our data twice up to lag 2, where we received a variance of 0.03016821. After attempting to difference the deseasonalized data once again at lag 2, the variance increased to 0.08036064. Because there is an increase in variance, the data is overdifferenced . Thus we do not need to difference again after the initial deseasonalizing and detrending.

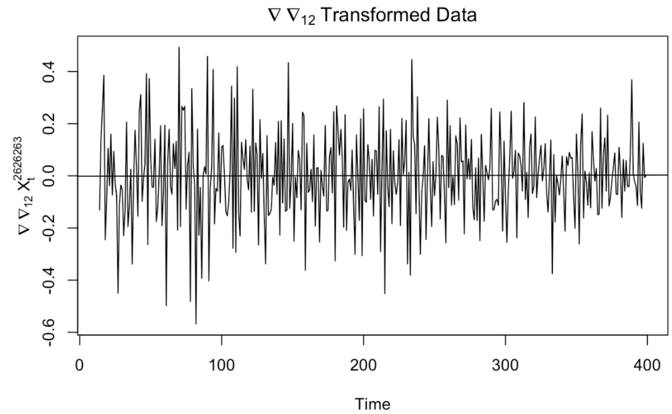


Figure 8: Graph of Detrended Data Post Second Differencing at Lag = 12

5 Model Identification

5.1 ACF and PACF Analysis

In figure 9 below, we first looked at the ACF and PACF at lags 12,24,36,... as our seasonal component is 12. The ACF cuts off after lag 12 indicating Q=1 or Q=0. The PACF tails off exponentially, so P=0 or P=1. In figure 10, we zoomed into the first plot to look at the ACF and PACF for lags less than 12. By looking at the second plot for ACF and PACF, we can see that the PACF cuts off after lag 2, indicating that P=2 or P=0 and ACF plot cuts off after 0, indicating that Q=0. In order to determine the specific model, we will test all combinations of P and Q from 0 to 2 and choose the two smallest AIC and BIC values for our models.

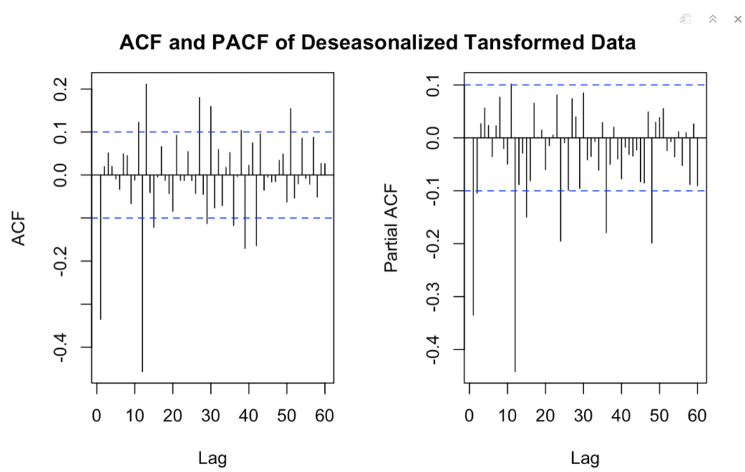


Figure 9: ACF and PACF of data post detrending and deseasonalization

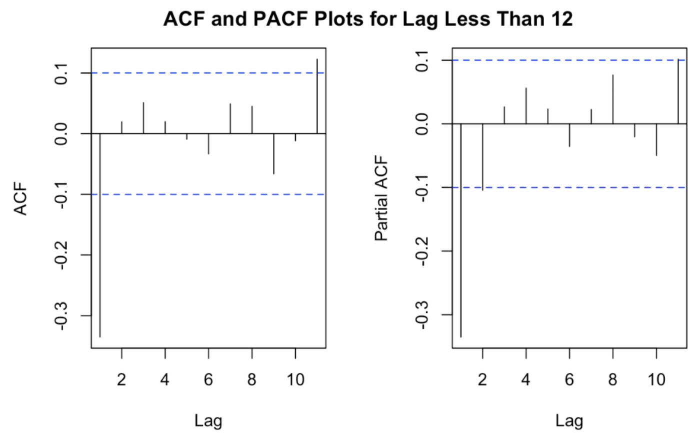


Figure 10: ACF and PACF of data post detrending and deseasonalization

5.2 Model Selection

From our analysis of the ACF and PACF, we can choose one of the models out our possible 9 models. In order to choose the proper model, we choose the model that has the minimum score for both BIC and AIC. As we can see from our tables, our models for the smallest AIC and BIC are the two

	q=0 <dbl>	q=1 <dbl>	q=2 <dbl>
p=0	−2.740412	−2.862373	−2.857273
p=1	−2.846063	−2.857267	−2.854452
p=2	−2.854470	−2.852342	−2.852633

Figure 11: AIC

	q=0 <dbl>	q=1 <dbl>	q=2 <dbl>
p=0	−3.735486	−3.847470	−3.832420
p=1	−3.831160	−3.832414	−3.819675
p=2	−3.829616	−3.817565	−3.807959

Figure 12: BIC

SARIMA models, model 1: $SARIMA(0,1,1)\text{x}(1,1,0)_{12}$ and model 2: $SARIMA(0,1,2)\text{x}(1,1,0)_{12}$. Using these two models, we can plot the roots of our two models, which can be seen on the following page, to check if they are causal and invertible.

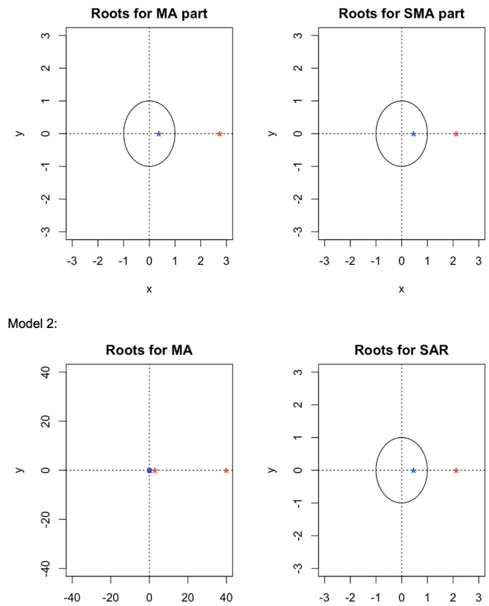


Figure 13: Roots for Causal and Invertibility Check

As we can see from our graphs, all of our roots lie outside the unit circle with the absolute value of the coefficients less than one. Therefore, model 1 and 2 are both causal and invertible.

6 Diagnostics

For our diagnostics of our two models, we use a significance level of .05. Our H_0 Hypothesis is that our residuals are normal, while the alternate hypothesis states that the residuals are not normal. Thus, if we get a p value greater than our alpha level of 0.5, we fail to reject H_0 and conclude our residuals are normal.

6.1 Shapiro Test

In order to diagnose our optimal model, we must check if one of the models does not pass the Shapiro test. The appendix includes more detailed information on this statistic and P value for each model.

	W.Statisttic <dbl>	P.value <dbl>
Model1	0.9959824	0.4135959
Model2	0.9959126	0.3977136

Figure 14: Shapiro Test Chart

6.2 Box Pierce and Ljung Box Test

In addition to the Shapiro Test, we shall conduct the Box Pierce and Ljung Box tests in order to check our P-Values pass our test.

	Model1.P.value <dbl>	Model2.P.value <dbl>
Box−Pierce	0.3752579	0.3737365
Ljung−Box	0.3531501	0.3516649

Figure 15: Box Pierce and Ljung Box Charts

6.3 Selecting Our Final Model

We can look at our possible two models, which have passed all of our tests, in Figure 18. From what we have shown above, our models have passed our Box Pierce, Ljung Box, and Shapiro tests. In addition, looking at the Figure in the appendix A.1, we can see that both models are normal based on our Q-Q plot and histogram. Thus, we choose the model with fewer coefficients, which is Model 1.

```
#model1, SARIMA(0,1,1)x(1,1,0)_12
#Fit and Estimation based on MLE method
Call:
arima(x = ts.data.minus.12, order = c(0, 1, 1), seasonal = list(order = c(1,
    1, 0), period = 12), method = "ML")

Coefficients:
          ma1     sar1
      -0.3645  -0.4720
s.e.   0.0463   0.0455

sigma^2 estimated as 0.0207:  log likelihood = 197.54,  aic = -391.08
```

```
#Model2: SARIMA(0,1,2)x(1,1,0)_12
Call:
arima(x = ts.data.minus.12, order = c(0, 1, 2), seasonal = list(order = c(1,
    1, 0), period = 12), method = "ML")

Coefficients:
          ma1     ma2     sar1
      -0.3685  0.0086  -0.4712
s.e.   0.0518  0.0486   0.0458

sigma^2 estimated as 0.0207:  log likelihood = 197.55,  aic = -389.11
```

Figure 16: Final Model Arima Call

Therefore, using X_t as the transformed of $X_t = \nabla\nabla_{12}Y_t^{.2626263}$, we can conclude our final model.

Final Model: SARIMA(0,1,1) x (1,1,0)$_{12}$

$$X_t = (1 - .3645B)(1 - .4720B^{12})Z_t$$

7 Forecasting

The importance of forecasting future data lies in the need to comprehend the future of the labor market for our youth. Using the 12 months we removed from our data, we are going to forecast a year's worth of data using our new, stationary model. We include general forecasting graphs below to see how our forecasting has been changed since transforming the data.

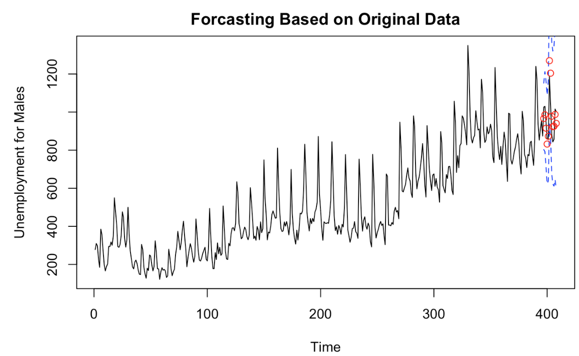


Figure 17: Forecast Using Original Data

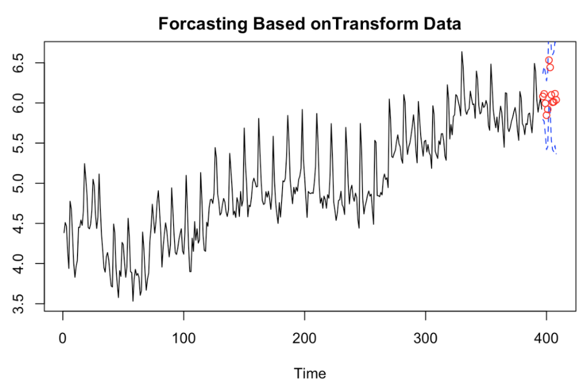


Figure 18: Forecast Using Transformed Data

In Figure 17, we forecasted 12 future values using our original, untransformed data. In Figure 18, we again forecasted 12 future values but used our transformed data. The red circles represent the forecasted values, while the blue dotted lines represent the bounds of confidence for our predictions. As we can see, our forecasted values fall inside the confidence bounds for the most part.

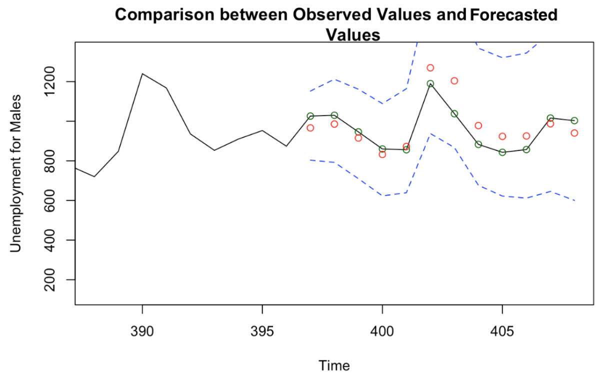


Figure 19: Forecast Comparison Between Transformed and Original Data

Figure 19 shows the comparison between our transformed predictions and our original data. The green dots indicate observed values and the red dots indicate forecasted values. We used the last 12 data points of the original data for our comparison. As the plot shows, forecasted data points 6 and 7, which represent the months of June and July, spike and then gradually fall. This aligns with our original observations in which we cited an increase in unemployment in the summer months of June and July. Our forecasted values also follow an upward trend and gradual fall towards the end of the year. It is evident that our forecasted values are close to the observed values of our original data and thus, we have chosen the correct model as our final model.

8 Conclusion

Our model has helped immensely in predicting the future of the labor market for males between the ages of 16 and 19. However, there are many future steps we could undertake in order to make this model even more useful.

One of the main issues with the model lies in the possibility of the market changing drastically which would distort the future unemployment rates. The distortion could come from a variety of different causes, such as a recession or depression, hyperinflation, or several other economic catastrophes which would ruin the labor market. One potential future study would be to utilize stationary time series models of recessions and depressions in order to relate our data to those trends and see how the labor market reacts.

In addition, we could compare the data to other age groups or a broader sample of the labor market. As many teenagers do not focus on jobs the way most adults do, the labor market could be drastically different compared to that of teenagers. For example, it's possible the seasonality could be different as adults look for work year round as they do not have summer breaks the way students do. We could also compare our data to more recent data to check if there has been a drastic change in the labor market over a span of time. This could allow us to see if the labor market is predictable over extremely large periods of time or if specific events completely throws it off, creating too unpredictable of a market.

To summarize our process, we needed to find a model which would help show the trend of unemployment in the labor market so that we may better understand what leads to problems in the labor market and how to prepare for these problems for the future. After finding our upward trend in unemployment over time as well as the seasonal increase in unemployment during the month of

June, we decided we needed to find a SARIMA model in order to make our time series stationary and allow for predictions of the future. In order to return this model, we went through the process of differencing and detrending our model in order to get a transformed data set. From here, we needed to select the proper model which best described our data. We analyzed the BIC and AIC values of our possible models and then utilized the Shapiro, Box Pierce, and Ljung Box tests in order to find our final best Model. Thus, we received our final model of

Final Model: SARIMA(0,1,1) x (1,1,0)$_{12}$

$$X_t = (1 - .3645B)(1 - .4720B^{12})Z_t$$
.

Below is the Appendix, R Code, and References including the data source we used in order to make this project possible. Thank you Professor Bapat for teaching us how to analyze and understand Time Series Models.

9 R Code

```
Appendix: R code

---
title: "Pstat 174 project"
author: "The Sudeep Sistas and the Bapat Bois"
date: "11/19/2018"
output: html_document
---

```{r}
plot.roots <- function(ar.roots=NULL, ma.roots=NULL, size=2, angles=FALSE, special=NULL,
    sqecial=NULL,my.pch=1,first.col="blue",second.col="red",main=NULL)
{xylims <- c(-size,size)
      omegas <- seq(0,2*pi,pi/500)
      temp <- exp(complex(real=rep(0,length(omegas)),imag=omegas))
      plot(Re(temp),Im(temp),typ="l",xlab="x",ylab="y",xlim=xylims,ylim=xylims,main=main)
      abline(v=0,lty="dotted")
      abline(h=0,lty="dotted")
      if(!is.null(ar.roots))
        {
          points(Re(1/ar.roots),Im(1/ar.roots),col=first.col,pch=my.pch)
          points(Re(ar.roots),Im(ar.roots),col=second.col,pch=my.pch)
        }
      if(!is.null(ma.roots))
        {
          points(Re(1/ma.roots),Im(1/ma.roots),pch="*",cex=1.5,col=first.col)
          points(Re(ma.roots),Im(ma.roots),pch="*",cex=1.5,col=second.col)
        }
      if(angles)
        {
          if(!is.null(ar.roots))
            {
              abline(a=0,b=Im(ar.roots[1])/Re(ar.roots[1]),lty="dotted")
              abline(a=0,b=Im(ar.roots[2])/Re(ar.roots[2]),lty="dotted")
            }
```
```

````
         if(!is.null(ma.roots))
           {
             sapply(1:length(ma.roots), function(j)
                 abline(a=0,b=Im(ma.roots[j])/Re(ma.roots[j]),lty="dotted"))
           }
       }
     if(!is.null(special))
       {
         lines(Re(special),Im(special),lwd=2)
       }
     if(!is.null(sqecial))
       {
         lines(Re(sqecial),Im(sqecial),lwd=2)
       }
       }
```

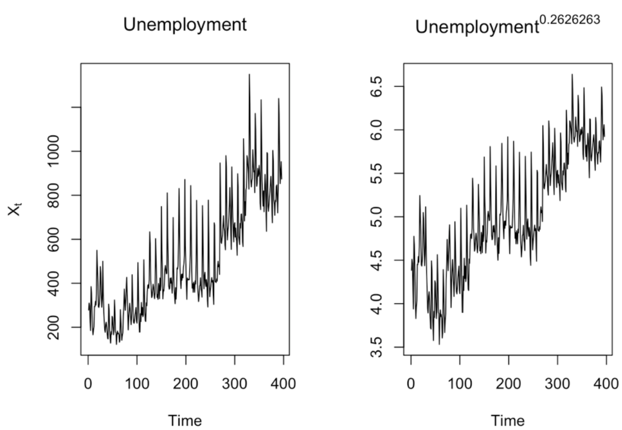
```{r}
library(MASS)
library(forecast)
library(ggplot2)
library(astsa)
library(tseries)
library(GeneCycle)
library(TSA)
```
```{r}
Data <- read.csv("Data.csv",sep=",", header=FALSE, skip=1)
data<-Data[1:408,]
head(data)
```

```{r}
ts.data <-ts(data[,2] ,start = c(1948,1),frequency=12)
plot(ts.data, xlab= "Time" , ylab="Unemployment for Males" , main= "Monthly T.S. Plot for
    Monthly U.S. Males (1948-1981) Unemployment")
```
```{r}
seasonplot(ts.data, 12, col=rainbow (3) , year.labels=TRUE, main= "", ylab="Unemployment
    for Males")
```

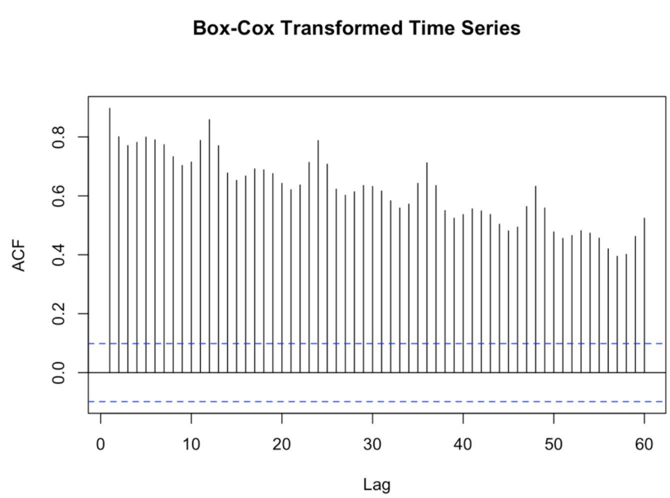


```{r}
decom <- decompose(ts.data)
autoplot(decom, main= "Decomposition Plot" ) +
theme ( axis.text.y = element_text(size =6), text = element_text(size=10)) +
xlab ( "Time in years" )
```
```{r}
t= 1:length(ts.data)
data.box = boxcox(ts.data ~ t ,plotit=TRUE)
lambda = data.box$x[which(data.box$y == max(data.box$y))] #0.2626263
ts.data.bc = ts.data^lambda
````

~~~
```
```{r}
var(ts.data) #68237.24
var(ts.data.bc)
```
```{r}
 op <- par(mfrow = c(1,2))
ts.plot(ts.data[1:(length(ts.data)-12)],main = expression(Unemployment),ylab =
    expression(X[t]))
ts.data.minus.12 <- ts(ts.data.bc[1:(length(ts.data)-12)]) #408-12=396
plot (ts.data.minus.12, xlab="Time", ylab="", main= expression( Unemployment^0.2626263))
```
~~~




~~~
```{r}
acf(ts.data.minus.12, lag.max = 60, main= "" )
title("Box-Cox Transformed Time Series", line = -1, outer=TRUE)
```
~~~



```
```{r}
data.diff.12 <- diff(ts.data.minus.12, lag=12)
plot(data.diff.12,xlab="Time",main = "De-seasonalized Time Series",
     ylab = expression(nabla~Transformed_Data))
abline(lm(data.diff.12~as.numeric(1:length(data.diff.12))))
var(data.diff.12) #0.07814674
```
```{r}
data.diff.12.diff.1 <- diff(data.diff.12, lag=1)
ts.plot(data.diff.12.diff.1,main = "De-trended/seasonalized Time Series",ylab
    =expression(nabla~nabla^{12}~Y[t]))
abline(lm(data.diff.12.diff.1~as.numeric(1:length(data.diff.12.diff.1))))
var(data.diff.12.diff.1) # 0.03020707
```
```{r}
data.diff.12.diff.2 <- diff(data.diff.12.diff.1, lag=1)
var(data.diff.12.diff.2) #0.08078615
```
```{r}
#using the Augmented Dickey-Fuller test to verify whether Xt is stationary or not.
adf.test(data.diff.12.diff.1)
```
```{r}
#identify P and Q
op<-par(mfrow=c(1,2))
acf(data.diff.12.diff.1,lag.max = 60,main="")
pacf(data.diff.12.diff.1,lag.max =60,main="")
title (main="ACF and PACF of Deseasonalized Transformed Data" ,outer=TRUE,line=-1)
```
```{r}
#identify p and q
#zoom acf and pacf plot
op<-par ( mfrow=c ( 1 , 2 ) )
acf(data.diff.12.diff.1 ,lag.max=11,main="")
pacf ( data.diff.12.diff.1 , lag.max=11,main="")
title (main="ACF and PACF Plots for Lag Less Than 12" ,outer=TRUE, line=-1)
par(op)
```
```{r}
#Model selection by AICc
AICc<-numeric()
for (p in 0:2){
  for (q in 0:2){
    AICc<-c(AICc,sarima(ts.data.minus.12,p,1,q,1,1,0,12,details = FALSE)$AICc)
  }
}
AICc<-matrix(AICc,nrow=3,byrow = TRUE)
rownames(AICc)<-c("p=0","p=1","p=2")
colnames(AICc)<-c("q=0","q=1","q=2")
AICc<-data.frame(AICc)
aicc<-setNames(AICc,c("q=0","q=1","q=2"))
aicc
```
```{r}
BIC<-numeric()
```
```

```
for(p in 0:2) {
  for (q in 0:2){
    BIC<-c(BIC,sarima(ts.data.minus.12,p,1,q,1,1,0,12,details = FALSE)$BIC)
  }
}
BIC<-matrix(BIC,nrow=3,byrow = TRUE)
rownames(BIC)<-c("p=0","p=1","p=2")
colnames(BIC)<-c("q=0","q=1","q=2")
BIC<-data.frame(BIC)
bic<-setNames(BIC,c("q=0","q=1","q=2"))
bic
```

```{r}
#Fit and Estimation based on MLE method
fit1 <-arima(ts.data.minus.12, order=c(0,1,1), seasonal=list(order=c(1,1,0),
    period=12),method="ML")
fit1
```

```
##
## Call:
## arima(x = ts.data.minus.12, order = c(0, 1, 1), seasonal = list(order = c(1,
##     1, 0), period = 12), method = "ML")
##
## Coefficients:
##          ma1     sar1
##      -0.3645  -0.4720
## s.e.  0.0463   0.0455
##
## sigma^2 estimated as 0.0207: log likelihood = 197.54, aic = -391.08
```

```{r}
par(mfrow = c(1,2))
plot.roots(NULL,polyroot(c(1,-0.3645 )),main="Roots for MA ", size = 3)
plot.roots(NULL,polyroot(c(1,-0.4720)),main="Roots for SMA", size = 3)
```

```{r}
fit2 <-arima(ts.data.minus.12, order=c(0,1,2), seasonal=list(order=c(1,1,0),
                                              period=12),method="ML")
fit2
```

```
##
## Call:
## arima(x = ts.data.minus.12, order = c(0, 1, 2), seasonal = list(order = c(1,
##     1, 0), period = 12), method = "ML")
##
## Coefficients:
##          ma1    ma2     sar1
##      -0.3685 0.0086  -0.4712
## s.e.  0.0518 0.0486   0.0458
##
## sigma^2 estimated as 0.0207: log likelihood = 197.55, aic = -389.11
```

```{r}
par(mfrow = c(1,2))
plot.roots(NULL,polyroot(c(1,-0.3685,0.0086 )),main="Roots for MA", size = 40)
plot.roots(NULL,polyroot(c(1,-0.4712)),main="Roots for SAR", size = 3)
```

~~~
```{r}
op<-par(mfrow=c(2,2))
resid1<-residuals(fit1)
resid2<-residuals(fit2)
hist(resid1 ,main="Histogram of Residuals") #good
qqnorm(resid1 , main="Normal QQ Plot for Model")
qqline(resid1)
hist(resid2 ,main="Histogram of Residuals") #good
qqnorm(resid2 , main="Normal QQ Plot for Model")
qqline(resid2)
```
~~~

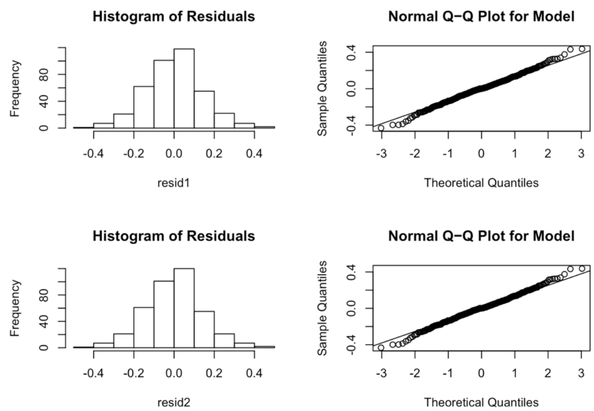


Figure A1: Histogram of Residuals and Normal QQ Plots

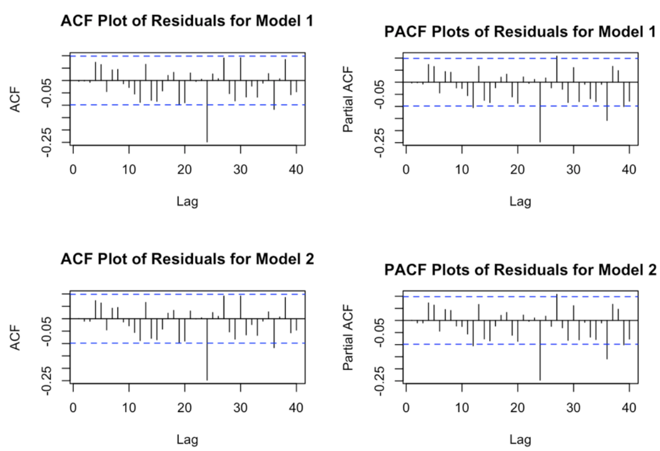
~~~
```{r}
#Shapiro Test for Model 1 and 2
Shap<-matrix(c(shapiro.test(resid1)$statistic
              ,shapiro.test(resid1)$p.value,
              shapiro.test(resid2)$statistic
              ,shapiro.test(resid2)$p.value),nrow=2,byrow=T)
#greater than 0.05 , then good
rownames(Shap)<-c("Model1" ,"Model2")
colnames(Shap)<-c("W Statisttic","Pvalue")
(Shap<-data.frame (Shap))
```

##        W.Statisttic  P.value
## Model1   0.9959824 0.4135959
## Model2   0.9959126 0.3977136
```{r}
b1<-Box.test(resid1, lag = 12, type = "Box-Pierce", fitdf = 2)$p.value
b2<-Box.test(resid1, lag = 12, type = "Ljung-Box", fitdf = 2)$p.value
b3<-Box.test(resid2, lag = 12, type = "Box-Pierce", fitdf = 2)$p.value
b4<-Box.test(resid2, lag = 12, type = "Ljung-Box", fitdf = 2)$p.value
boxT<-matrix(c(b1,b2,b3,b4) ,nrow=2,byrow=FALSE)
colnames(boxT)<-c("BoxPierce"," LjungBox ")
rownames(boxT)<-c("Model1 Pvalue" , "Model2 Pvalue")
(boxT<-data.frame(boxT))
```

##                Box.Pierce Ljung.Box
~~~

````
## Model1  Pvalue   0.3752579 0.3737365
## Model2  Pvalue   0.3531501 0.3516649
```{r}
#Test for constant variance of residuals
par(mfrow=c(2 ,2) ) # acf
acf(resid1, main = "ACF Plot of Residuals for Model 1" , lag.max=40) # pacf
pacf(resid1,main="" , lag.max=40)
title(main="PACF Plots of Residuals for Model 1",outer=FALSE,line=1) # acf

acf(resid2, main = "ACF Plot of Residuals for Model 2" , lag.max=40) # pacf
pacf(resid2,main="" , lag.max=40)
title(main="PACF Plots of Residuals for Model 2",outer=FALSE,line=1) # acf
```
````




````
```{r}
pred.tr <-predict(fit1 ,n.ahead = 12)
U.tr= pred.tr$pred+2*pred.tr$se # upper bound for the C. I . for transformed data
L.tr= pred.tr$pred-2*pred.tr$se # lower bound
ts.plot(ts.data.minus.12,xlim=c(1,length(ts.data.minus.12)+12), main="Forcasting Based on
    Transform Data",ylab="")
lines(U.tr,col="blue", lty="dashed")
lines(L.tr,col="blue", lty="dashed")
points((length(ts.data.minus.12)+1):(length(ts.data.minus.12)+12),pred.tr$pred, col="red")
```
```{r}
pred.orig<-pred.tr$pred^(1/lambda)
#  backtransform  to get predictions of original time series
U= U.tr^(1/lambda) # bounds of the confidence intervals
L=L.tr^(1/lambda)
# Plot forecasts with original data
ts.data2<-ts(data[ , 2 ] )
ts.plot(ts.data2 , xlim=c(1,length(ts.data2)) ,main="Forcasting Based on Original
    Data",ylab="")
lines(U, col="blue", lty="dashed")
lines(L, col="blue", lty="dashed")
points((length(ts.data.minus.12)+1):(length(ts.data.minus.12)+12), pred.orig ,col="red")
````
````

```
'''
'''{r}
#zoom effect
ts.plot(ts.data2 ,xlim=c(length(ts.data2)-20,length(ts.data2)), main="Comparison between
    Observed Values and Forecasted Values",ylab="")
points((length(ts.data.minus.12)+1):(length(ts.data.minus.12)+12), ts.data2[397:408],
    col="dark green")
points((length(ts.data.minus.12)+1):(length(ts.data.minus.12)+12),pred.orig, col="red")
lines((length(ts.data.minus.12)+1):(length(ts.data.minus.12)+12),U, lty=2, col="blue")
lines((length(ts.data.minus.12)+1):(length(ts.data.minus.12)+12),L, lty=2, col="blue")
'''
```

# References

[1] Sudeep Bapat: PSTAT174 Lecture Series Fall 2018,
https://gauchospace.ucsb.edu/courses/course/view.php?id=34152

[2] DataMarket Unemployment Data,
https://datamarket.com/data/set/22q4/monthly-us-male-16-19-years-unemployment-figures-thousands-1948-1981!ds=22q4display=line