US Unemployment Trends

Initial Model Selection

Group 4

Joseph Blubaugh Sean Roberson Akarshan Puri
Alison Shelton Travis Lilley Bo Pang

STAT 626: Time Series Analysis

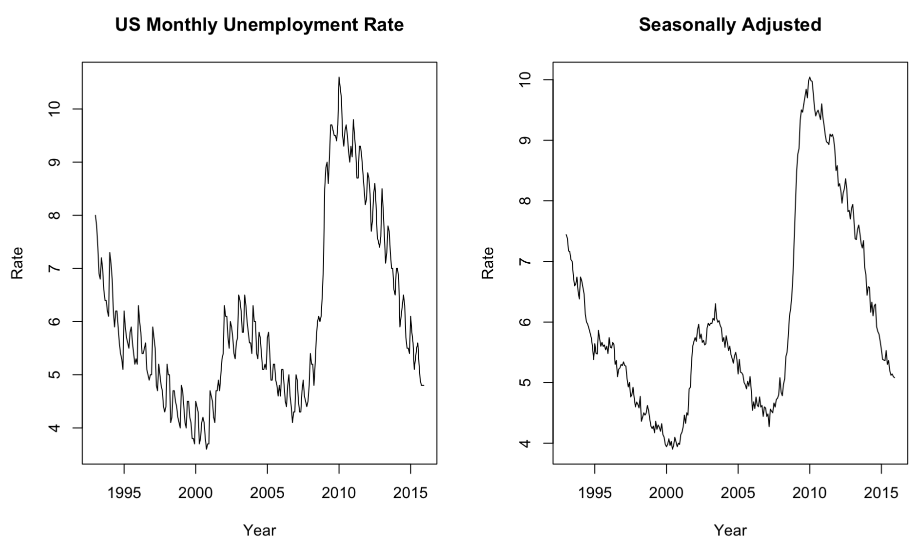
Introduction

Unemployment has been a topic of concern throughout the United States in recent years. Graduate and Undergraduate college students alike are concerned over their employment prospects, wondering if their degrees will be enough to gain them a job after graduation. In these times of economic uncertainty, obtaining an income generating position is not the guarantee it has seemed to be in generations past. Therefore, the purpose of our project is to examine trends in unemployment in the United States, focusing on the years from 1992 to 2015, with a goal of forecasting into late 2016 and beyond.

The unemployment data being examined was obtained from the non-seasonaly adjusted, monthly, Civilian Unemployment Rate Series (UNRATENSA). In this U.S. Bureau of Labor Statistics (BLS) and included figures from January of 1948 to May of 2016 [U.S. Bureau of Labor Statistics, 2016]. The response variable being analyzed is the unemployment rate defined as the percentage of the labor force that is unemployed. In defining this variable, the BLS restricts this to, "people 16 years of age and older, who currently reside in 1 of the 50 states or the District of Columbia, who do not reside in institutions (e.g., penal and mental facilities, homes for the aged), and who are not on active duty in the Armed Forces".

As a first step, the data was plotted over time to identify any obvious patterns visually, considering both seasonlly adjusted and non-seasonlly adjusted versions of the unemployment rate, See Figure 1. The time span included in the study encompases the presidential terms of Bill Clinton, George W. Bush, and Barack Obama, each serving eight years in office. Initial graphs of the data seem to indicate that, in general, unemployment spiked at the begining of each president's term and fell gradually over the time he was in office. There are also two noticeable spikes the represent that recessions of 2001 and 2008, respectively. The 2008 recession also follows the burst of a housing market bubble. These are all potential explanatory variables that will be explored in further analysis.

Figure 1: Plot of the original data



Stationarity

After the initial exploration of the time series graphs, the team has chosen to focus on building priliminary models which will serve as the foundation of further analysis. In looking at the inital plots, it appears that the series could benefit from detrending. As such, one of the primary goals has been to transform the data to stationarity.

An Augmented Dickey-Fuller (ADF) test for stationarity was conducted to verify the nonstationarity of the unemployment data. The ADF test tests the null hypothesis that the time series data has a unit root against the alternative that the data are stationary [Shumway and Stoffer, 2010]. For the non-seasonlly adjusted data the Dickey-Fuller test statistic was -1.4266, a lag order of 6, and a p-value of 0.8176. Additionally, the seasonally adjusted data had a Dickey-Fuller test statistic of -2.1377, a lag order of 6, and a p-value of 0.518. The high p-values suggest that we do not have a stationary model with just the raw unemployment data, whether or not the data are seasonally adjusted.

The first differences of the unemployment data were plotted for both the seasonal and seasonally adjusted unemployment data, see Figures 2 and 3. Both the first and second differences appear considerably more stationary when compared to the original data. The associated ADF test results are given in 1. Based on the p-values, there is significant evidence of stationarilty with each of the 4 differenced models. The consensus in the group was to continue the model building using second differences.

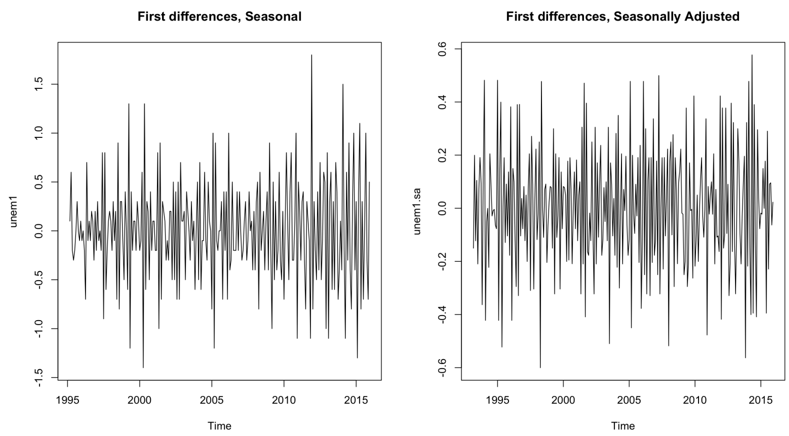


Figure 2: Plots of first differences

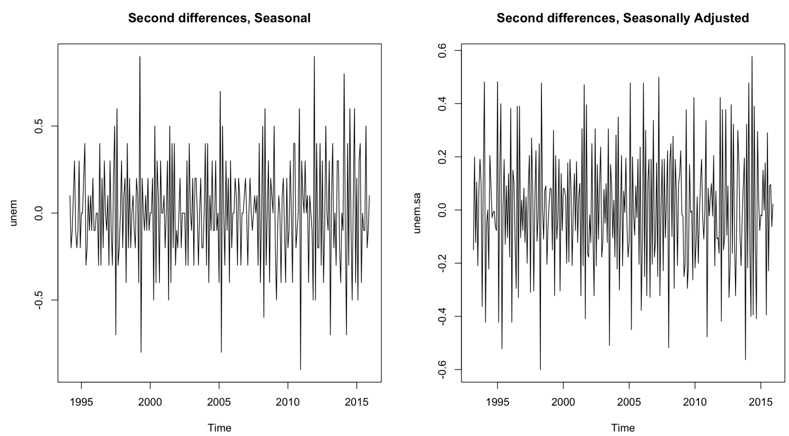


Figure 3: Plots of second differences

Table 1: ADF Test Results

Model	Statistic	Lag order	p-value
1^{st} difference	-7.6799	6	< 0.01
1^{st} difference, SA	-9.3595	6	< 0.01
2^{nd} difference	-8.4515	6	< 0.01
2^{nd} difference, SA	-9.3595	6	< 0.01

Building the ARIMA & SARIMA models

The team visually analyzed the ACF and PACF plots within the first season (h = 1, 2, ..., 12), see Figure 4. The PACF appears to decline slowly, while the ACF seems to fall off after 1. Therefore we began by letting p = 0, and q

= 1. Several models were considered by making adjustments to variations resulting in the models found in Table 2.

Figure 4: ACF & PACF Plots

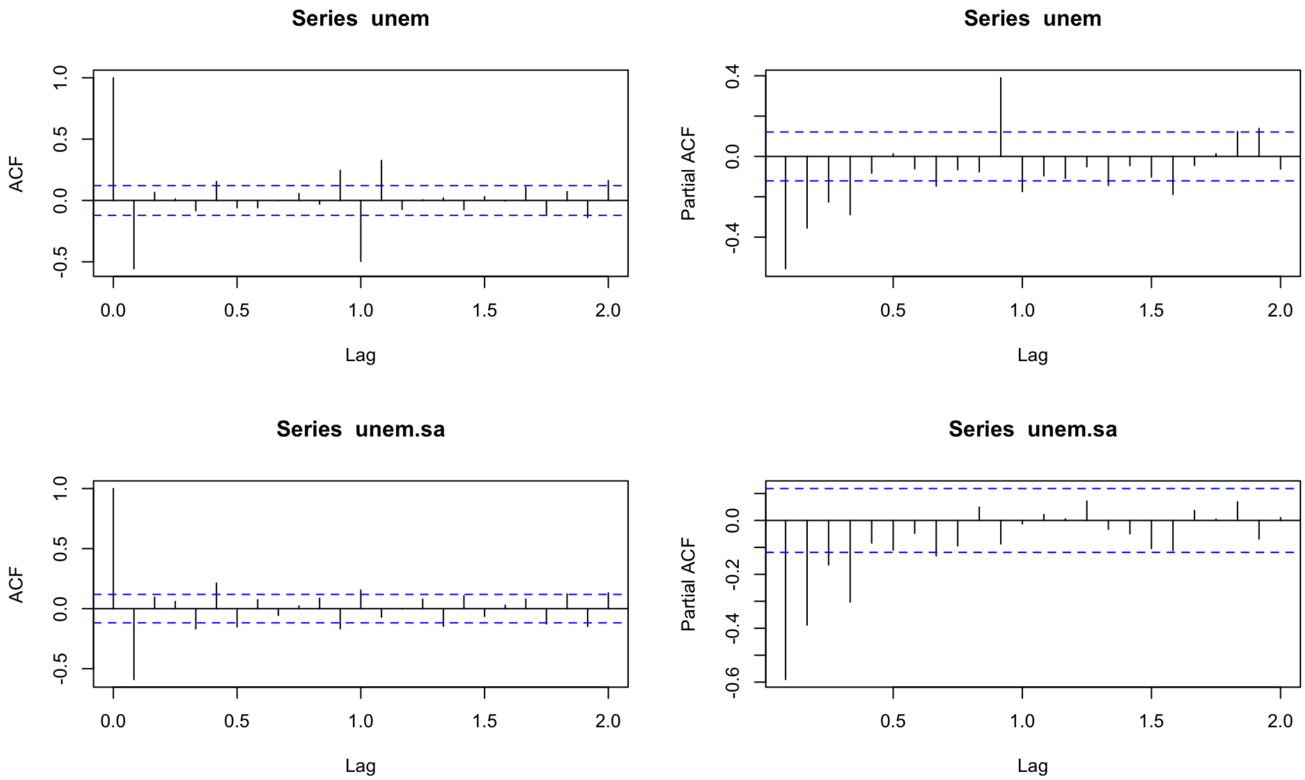


Table 2: Model Summaries

#	Data	Order	Seasonal Order	XRegs	AIC	BIC
1	Unem	0,2,1	1,1,0	N	-2.27	-3.23
2	Unem	0,2,1	3,1,0	N	-2.44	-3.37
3	Unem	4,2,1	3,1,0	N	-2.44	-3.32
4	Unem.sa	0,2,1	1,0,0	N	-2.61	-3.58
5	Unem.sa	1,2,1		N	-2.63	-3.60
6	Unem.sa	1,0,1		Y	-2.617	-3.578
7	Unem.sa	1,0,1		Y	-2.672	-3.565

Based on the AIC values the two models that show the most promise are models 5 and 7. Model 5 includes only the time series data whereas model 7 also includes some of the predictors of interest. The diagnostic plots are shown in Figures 5 and 6. Both models show a great deal of promise. The standardized residuals show no apparent pattern. The ACF of the residuals show no departure from normality. Although the Normal Q-Q plot of the standardized residuals shows some slight departure from normality in the tails, there is no strong evidence of lack of normality in the residuals The p-values for the Ljung-Box statistic are high enough at all plotted lags, there is no indication of lack of fit in the models. Therefore, we will continue to refine these models further as we explore the nature of US Unemployment rate patterns.

Figure 5: Model 5 Diagnostics

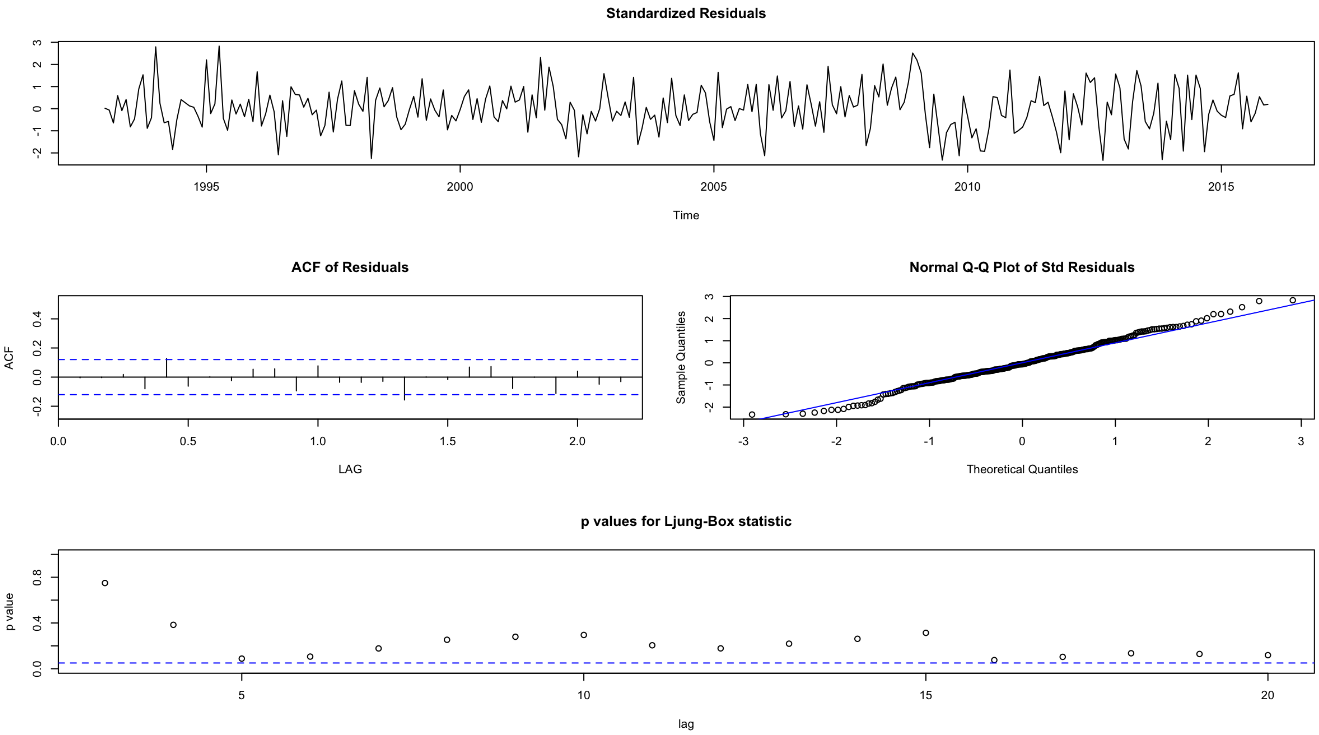
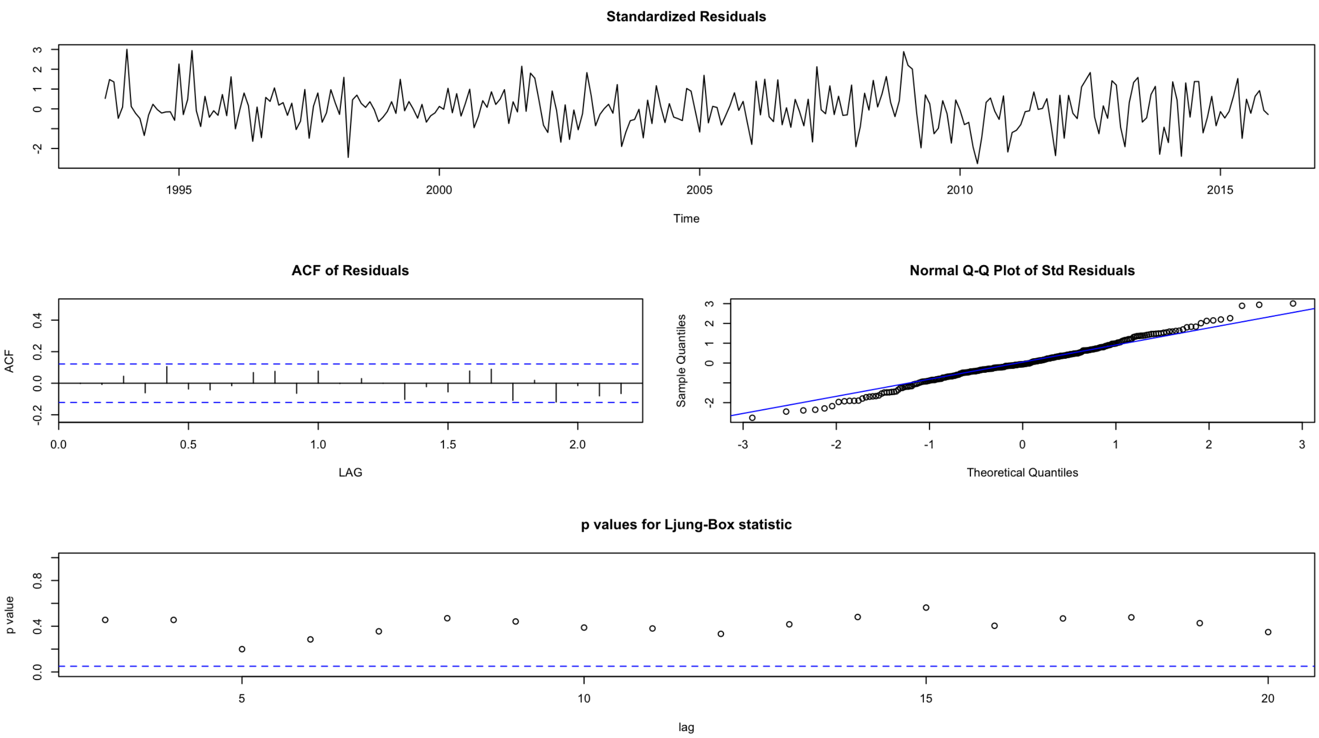


Figure 6: Model 7 Diagnostics



References

Robert H Shumway and David S Stoffer. *Time series analysis and its applications: with R examples.* Springer Science & Business Media, 2010.

U.S. Bureau of Labor Statistics. *Civilian Unemployment Rate [UNRATE].* https://fred.stlouisfed.org/series/UNRATE, 2016.