US Unemployment Trends

Initial Model Selection

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

STAT 626: Time Series Analysis

The following is just a collection of the discussions in preparation for turning what we have done into a write-up. This is not an actual write-up.

1 Achieving Stationarity

1.1 Testing for Stationarity

- I performed the ADF (augmented Dickey-Fuller) test for stationarity. We haven't talked about it in class yet, but it seems to be used pretty commonly. The high p-value suggests that we do not have a stationary model with just the raw unemployment data. This is confirmed with the non-constant variance and time-dependent mean that we observed in our exploratory analysis.
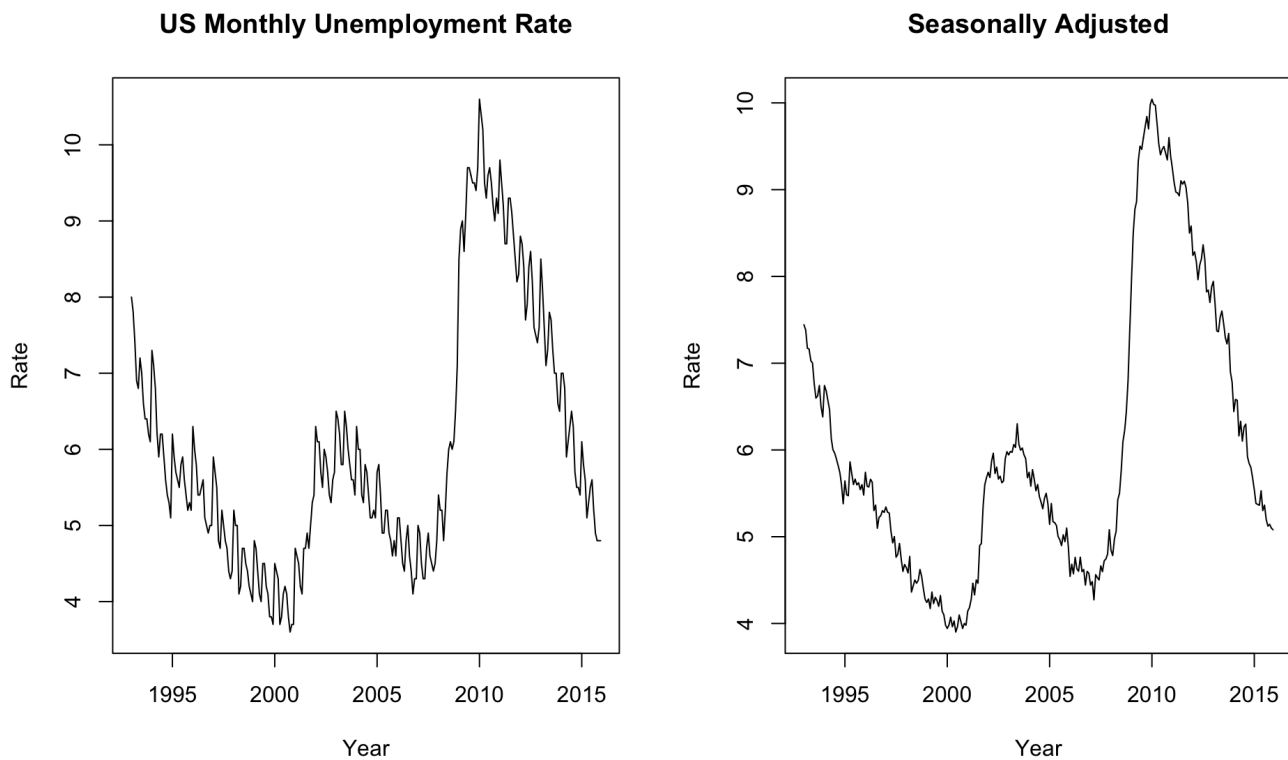
Augmented Dickey-Fuller Test

data: unem
Dickey-Fuller = -1.4266, Lag order = 6,
p-value = 0.8176
alternative hypothesis: stationary

data: unem.sa
Dickey-Fuller = -2.1377, Lag order = 6,
p-value = 0.518
alternative hypothesis: stationary

1.2 Initial Plots

Figure 1: Plot of the original data

I then plotted the first difference, and it appears relatively stationary.

This is confirmed by the ADF test on the first differenced data, which has a very small p-value:

Augmented Dickey-Fuller Test
Dickey-Fuller = -4.3501, Lag order = 6, p-value = 0.01
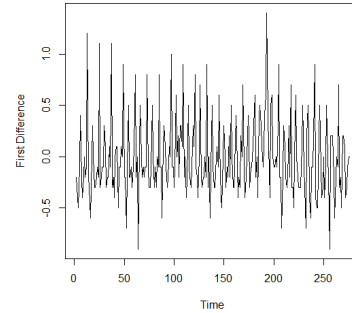alternative hypothesis: stationary


Figure 2: Plot of first Difference

1.3 Differencing

I took the second difference (d = 2), as Joseph suggested, then the first seasonal difference (D = 1) with s = 12 (this is common for monthly economic data), as the book did.

Below, Figure 4, is the plot of the transformed graphed. It looks pretty stationary (not perfect, but adequate), and we can confirm this with the ADF test (it's cited in other time series texts, but I haven't seen it in ours yet).

I like the idea of using the ADF. We can include it as part of our literature review.

Figure 3: Second differences with and without seasonal adjustments
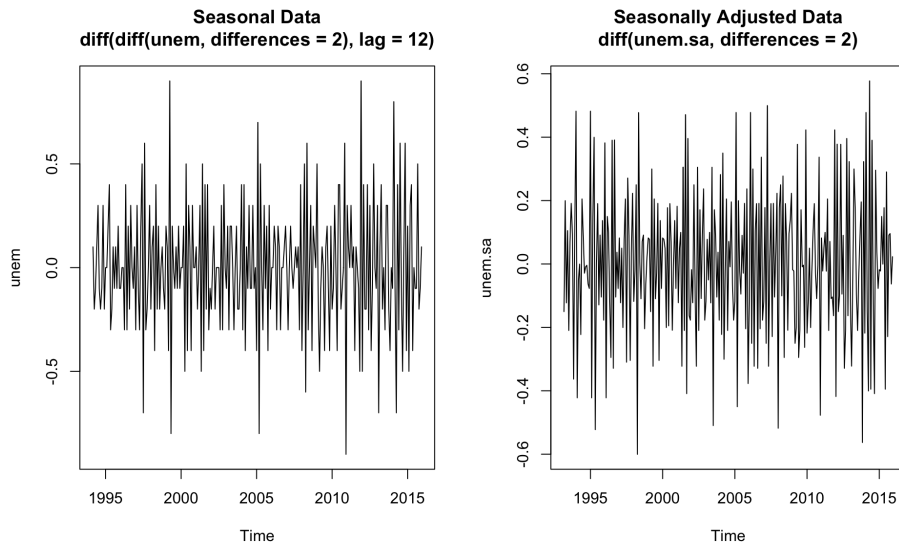

1.4 ACF & PACF

After that, the book suggests that you examine the ACF and PACF plots.

First, the book says to look at the seasonal changes in ACF and PACF (h = 12, 24, 36, ...). These seem to indicate that the ACF trails off, and the PACF cuts off after one year (h = 12). This suggests that we let P = 1 and Q = 0.

Next, the book says to look at the ACF and PACF within only the first season (h = 1, 2, ..., 12). The PACF declines slowly, but the ACF cuts off after 1, suggesting we let p = 0, and q = 1.

Figure 4: ACF & PACF Plots

Series unem

Series unem

Series unem.sa

Series unem.sa
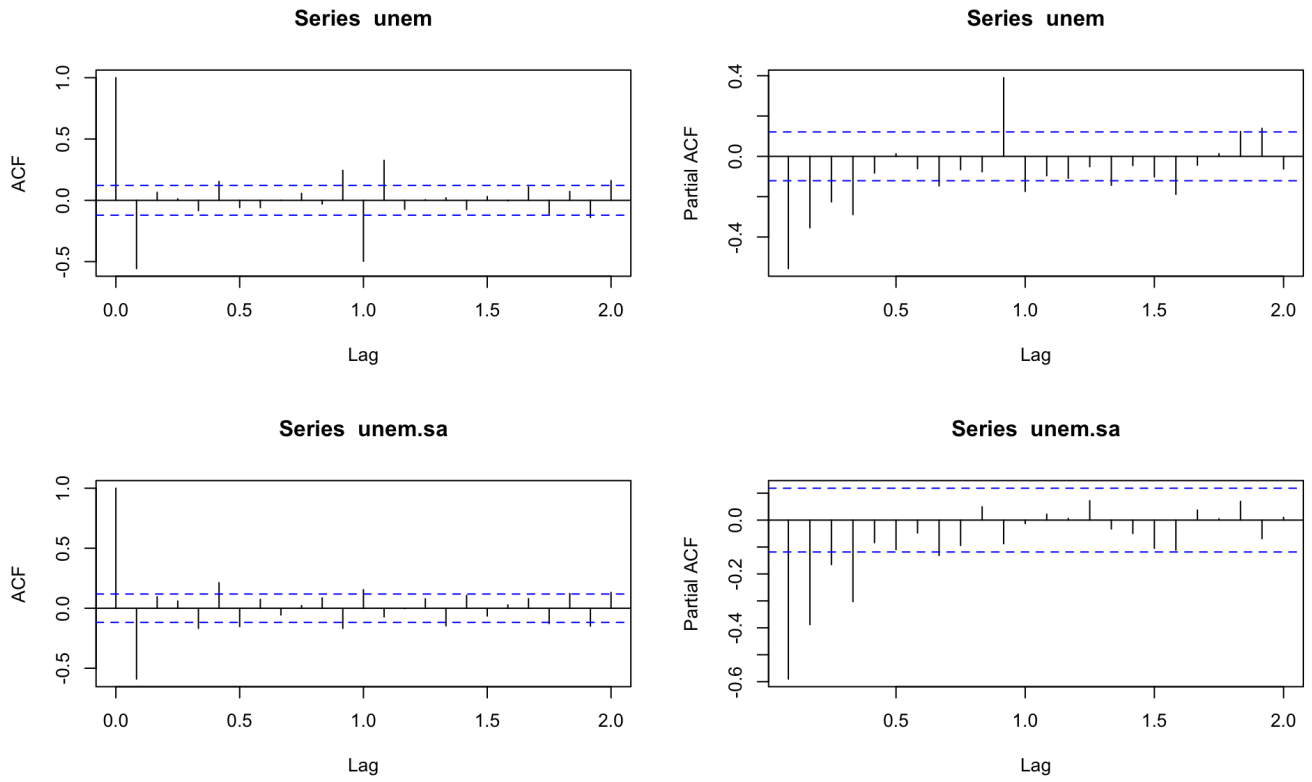
2 Model Comparisons

2.1 Overview

Models Considered

Table 1: Model Summaries

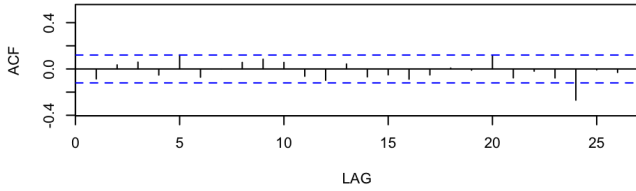| #	Data	Order	Seasonal Order	XRegs	AIC	BIC
1	Unem	0,2,1	1,1,0	N	-2.27	-3.23
2	Unem	0,2,1	3,1,0	N	-2.44	-3.37
3	Unem	4,2,1	3,1,0	N	-2.44	-3.32
4	Unem.sa	0,2,1	1,0,0	N	-2.61	-3.58
5	Unem.sa	1,2,1		N	-2.63	-3.60
6	Unem.sa	0,2,1	1,0,0	Y	-2.58	-3.49
7	Unem.sa	1,2,1		Y	-2.60	-3.49

2.2 Seasonal Models

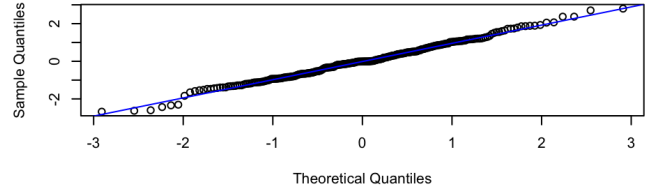**Model 1: SARIMA**$(0, 2, 1) \times (1, 1, 0)_{12}$
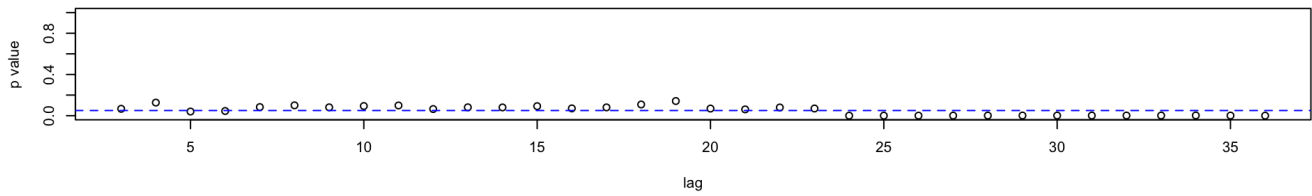
Standardized Residuals
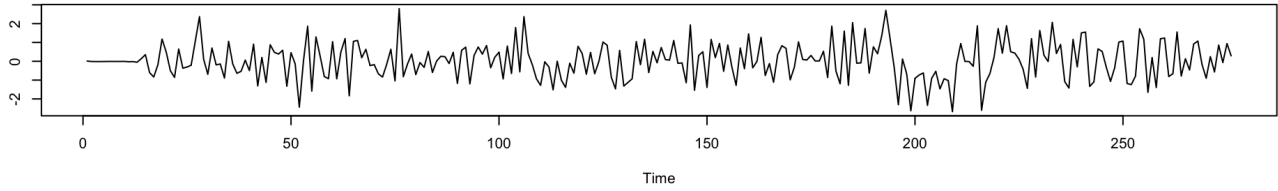
ACF of Residuals
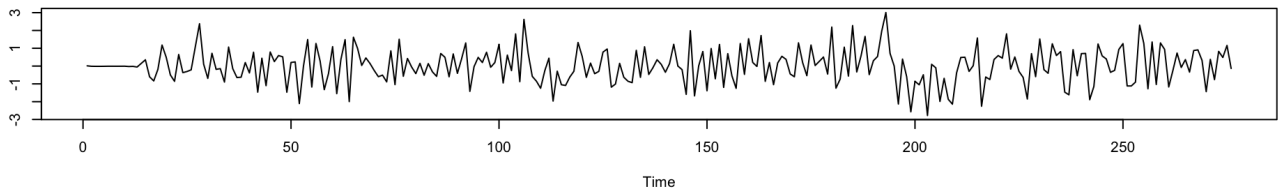

Normal Q-Q Plot of Std Residuals

p values for Ljung-Box statistic

Model 2: SARIMA$(0,2,1) \times (3,1,0)_{12}$

Standardized Residuals

ACF of Residuals

Normal Q-Q Plot of Std Residuals

p values for Ljung-Box statistic

Model 3: SARIMA$(4,2,1) \times (3,1,0)_{12}$

Standardized Residuals

ACF of Residuals

Normal Q-Q Plot of Std Residuals
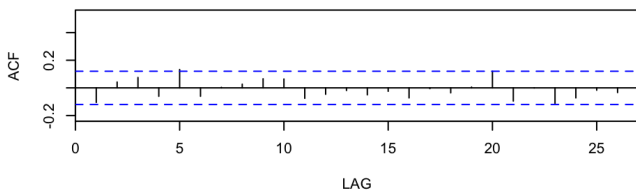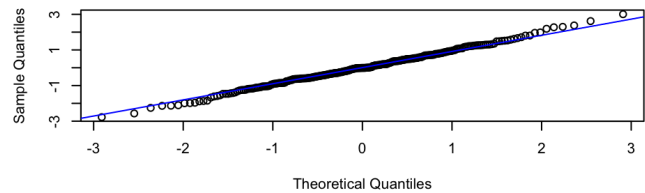

p values for Ljung-Box statistic
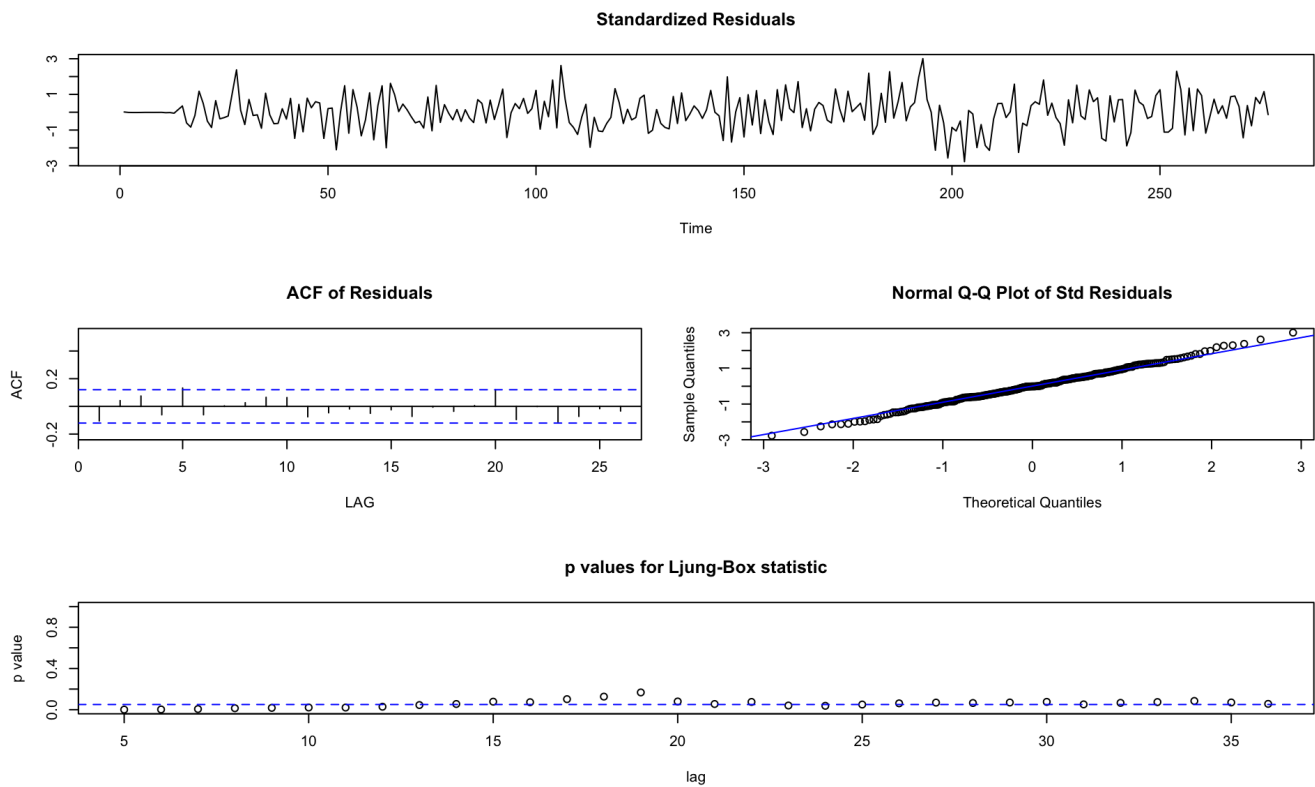
2.3   Seasonally Adjusted Models

Model 4: SARIMA$(0, 2, 1) \times (1, 0, 0)_{12}$

Standardized Residuals

ACF of Residuals

Normal Q-Q Plot of Std Residuals

p values for Ljung-Box statistic
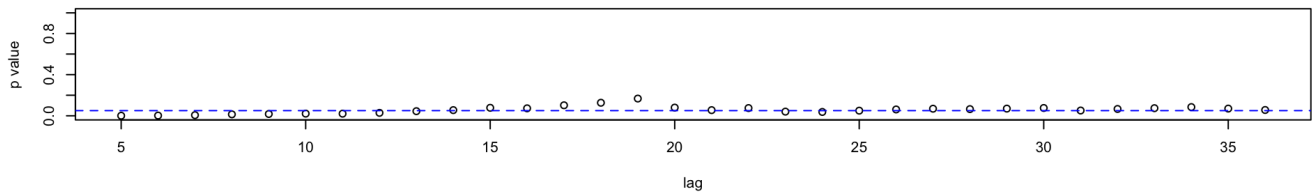
Model 5: ARIMA$(1, 2, 1)$

Standardized Residuals

ACF of Residuals

Normal Q-Q Plot of Std Residuals

p values for Ljung-Box statistic

Model 6: SARIMA$(0, 2, 1) \times (1, 0, 0)_{12}$ with Regressors

Standardized Residuals

ACF of Residuals

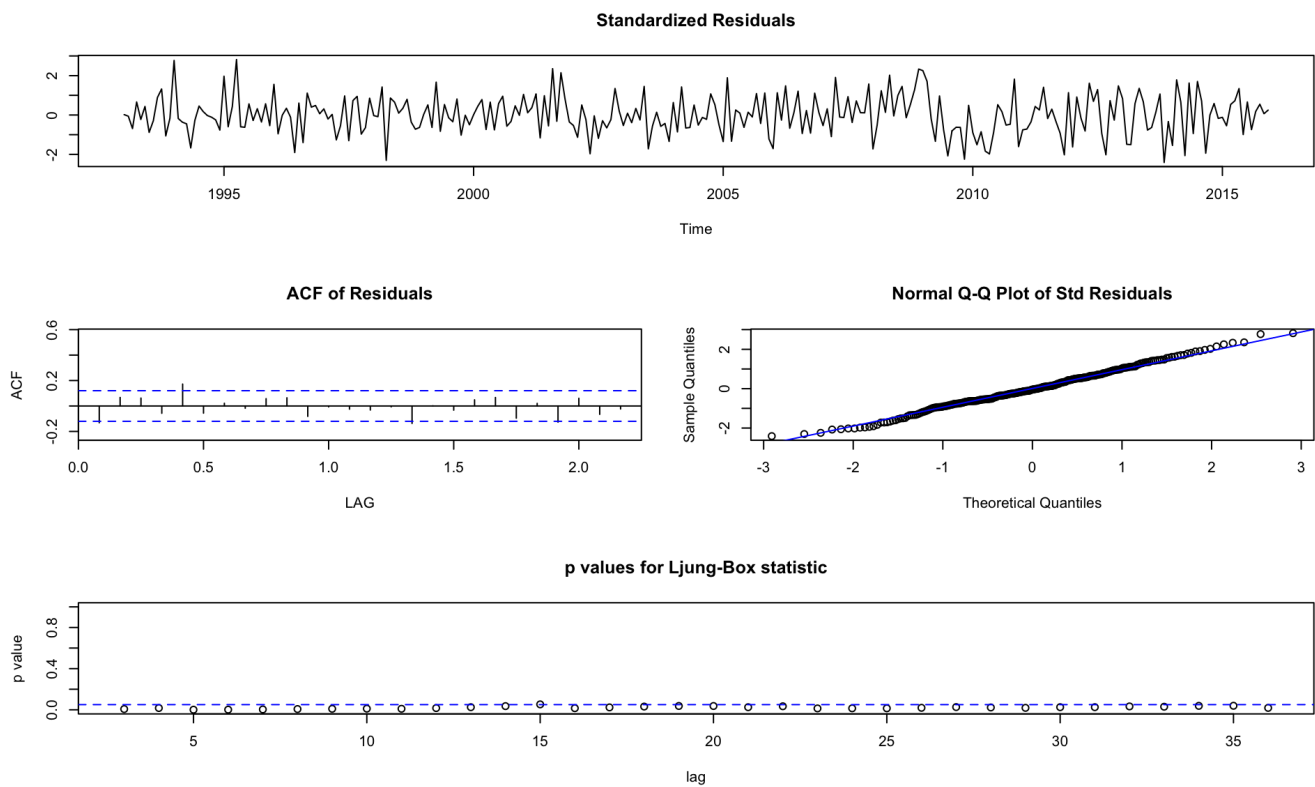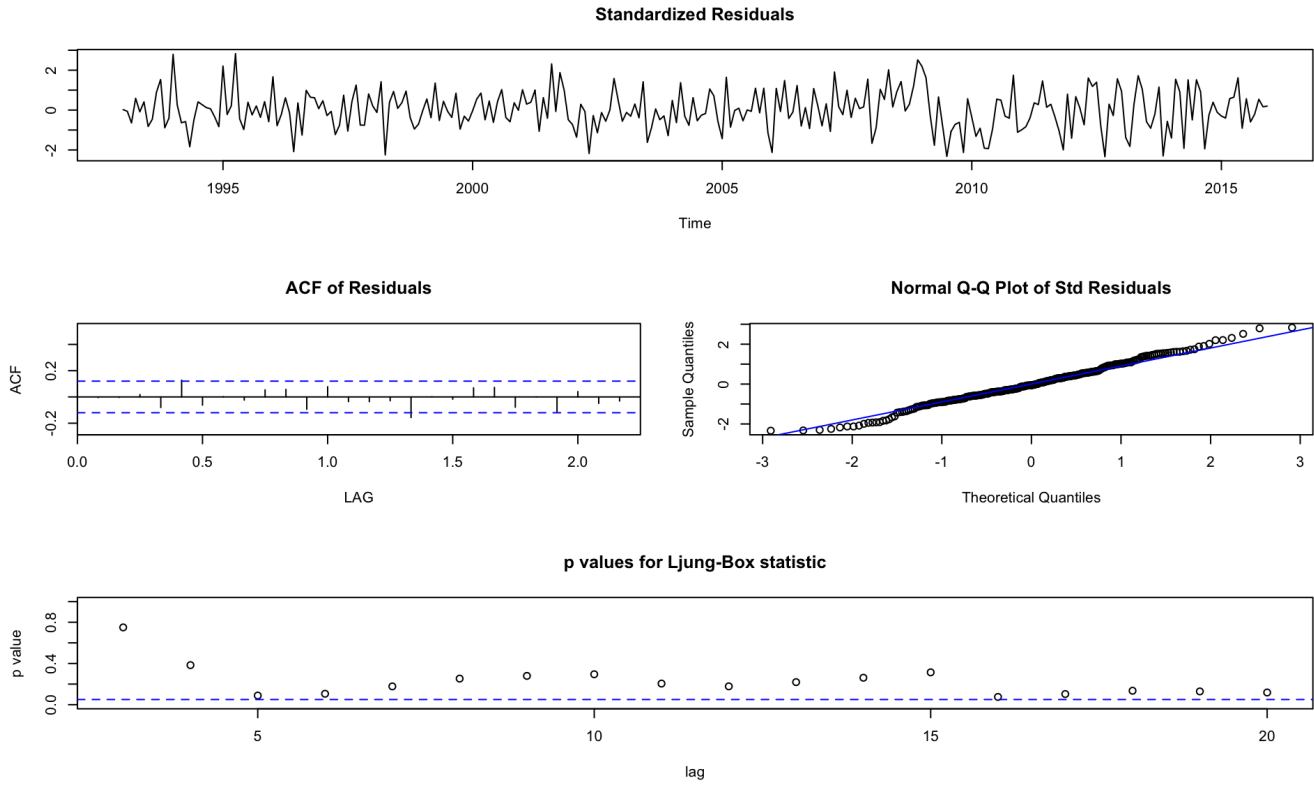**Normal Q-Q Plot of Std Residuals**

p values for Ljung-Box statistic
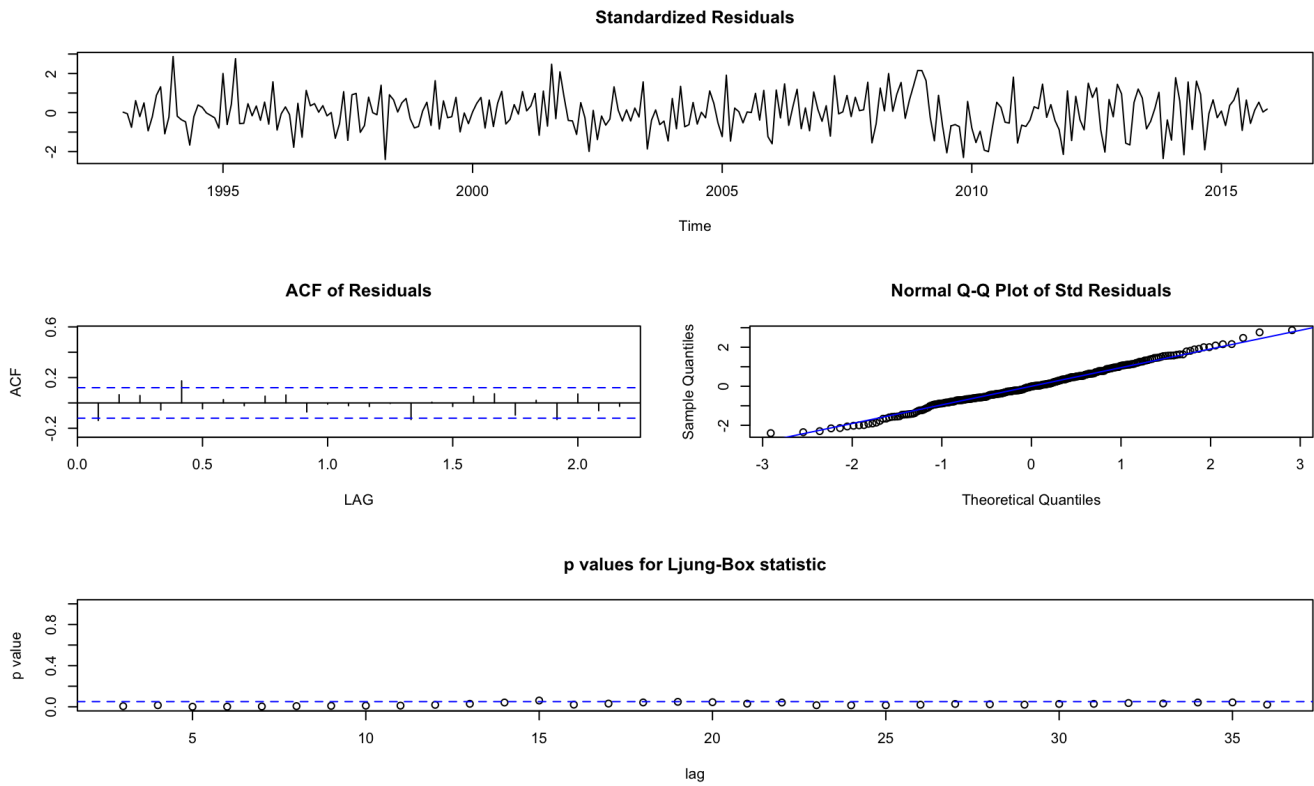
Model 7: ARIMA$(1, 2, 1)$ with Regressors

Standardized Residuals

ACF of Residuals

Normal Q-Q Plot of Std Residuals

p values for Ljung-Box statistic
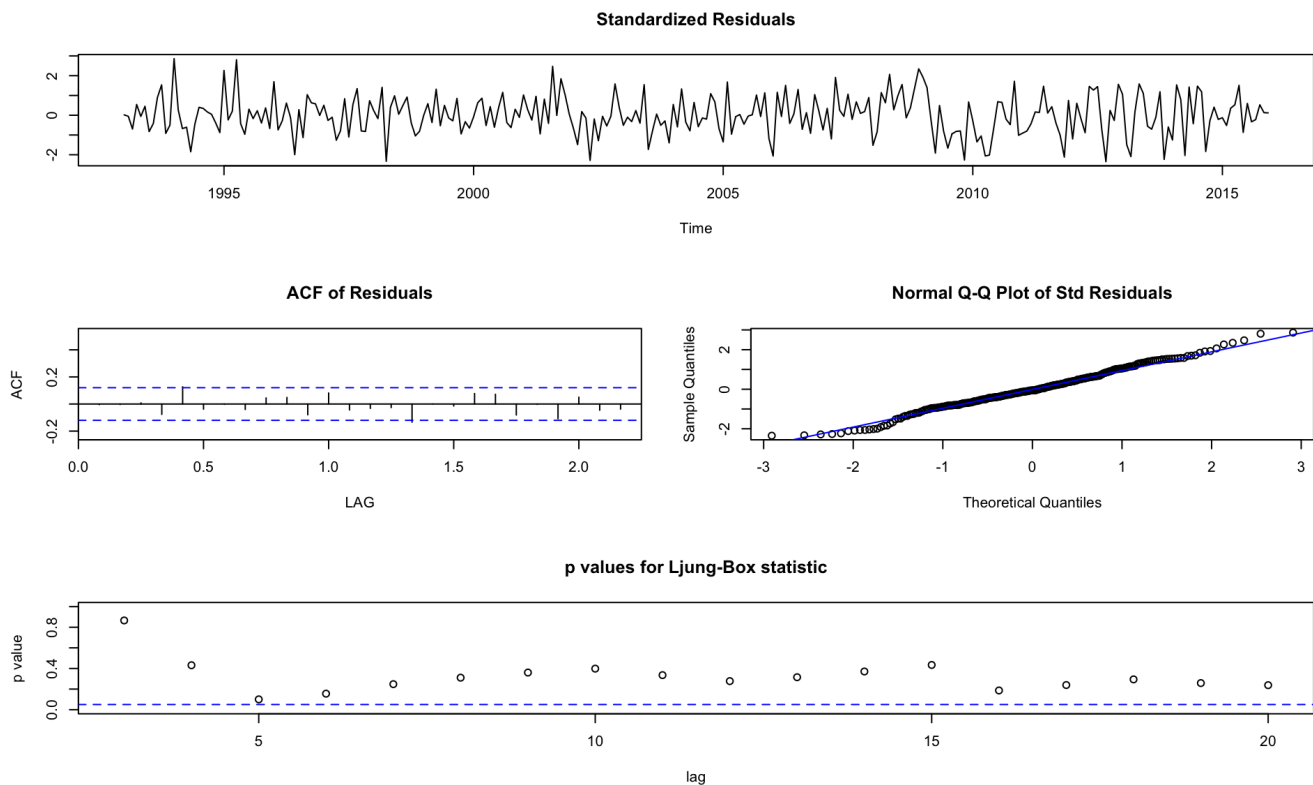
From looking at AIC and BIC values, Models 2 and 3 perform quite similarly, which both show some slight evidence of outperforming Model1.

Then we could compare the three models based on diagnostic plots. The standardized residuals of all models show some evidence of non-white-noise. ARMA models do not model variability. We will have a few lectures on this topic. There is not much we can do now on this issue.

In the ACF of residuals of Model 1 shows a spike at lag 24. The other two models do not show such a spike.

The normal plots from the three models are fairly similar.

The Q-statistic or Ljung-Box statistic

Models 1 and 2 have similar results. Model 1 seems to perform better at the first few lags, but Model 2 does better after lag 15. Model 3 clearly perform better than the two models on the Q-statistic. Since the Model 3 is based on a reasoning our professor does not like, we may not present this model. However it at least informs us that some models based on the thought that both the ACF and PACF cuts off at certain lags might model our data better.