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.
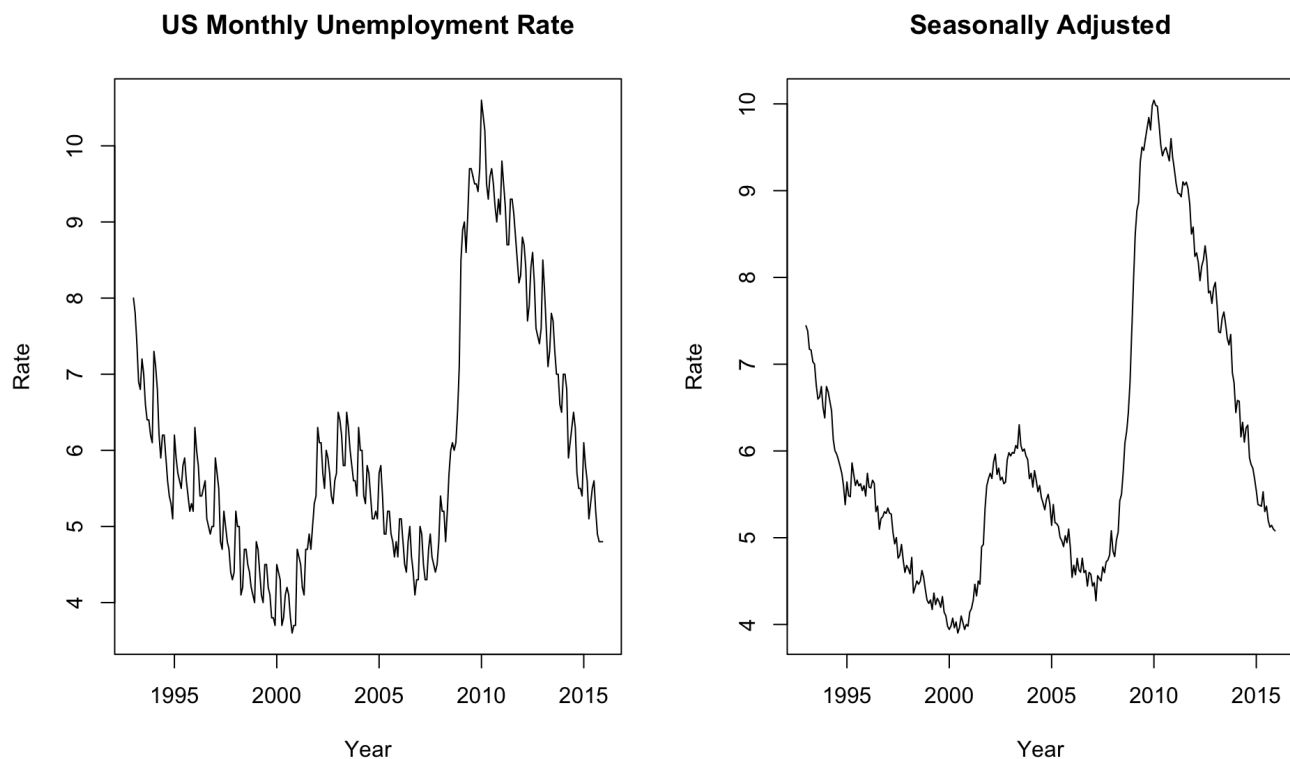
Augmented Dickey-Fuller Test

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
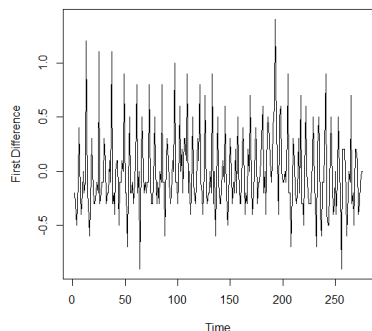
Figure 2: Plot of first Difference

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 3, 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
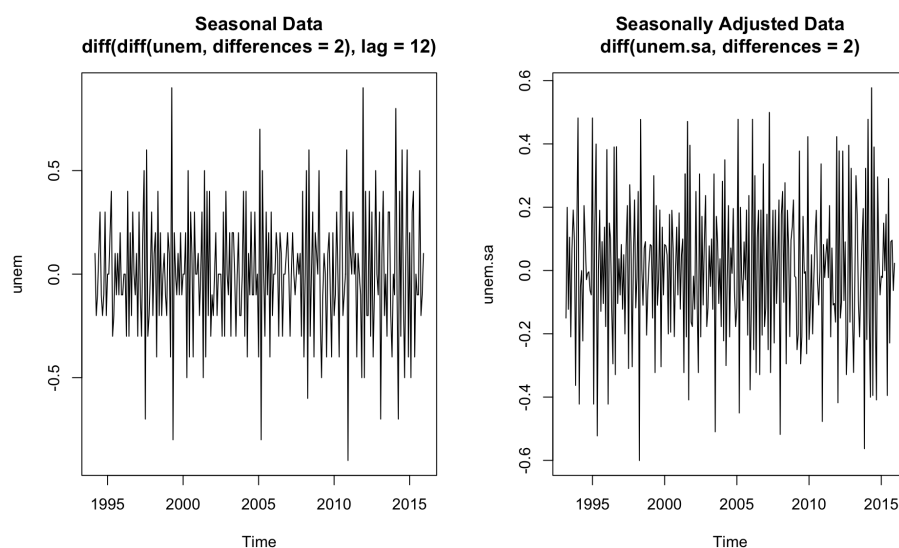
1.4 ACF & PACF

- 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.
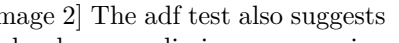
Figure 4: ACF & PACF Plots
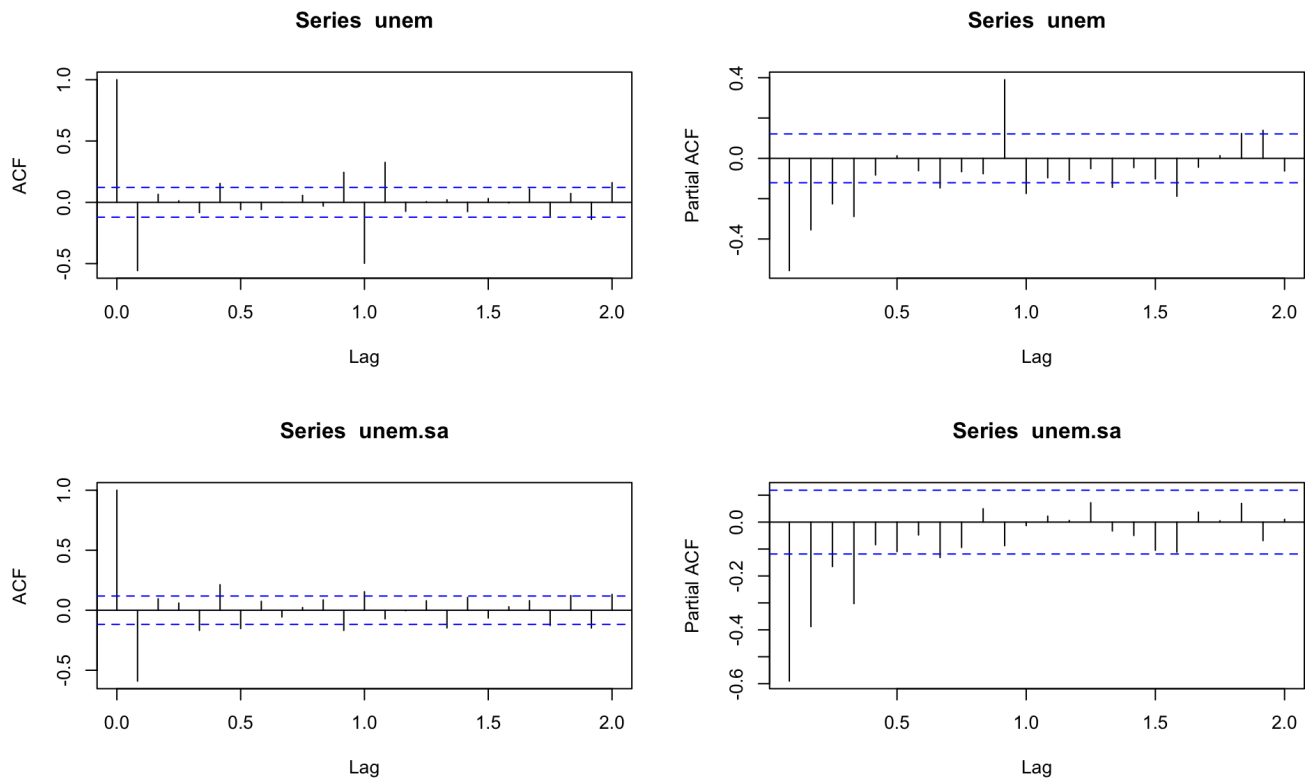
2 Model Comparisons

- I then look at the ACF and PACF plots to build models. First, we can look into the seasonal pattern. The ACF seems to tail off and the PACF cuts off at either 1 or 3. Together, the plots suggest AR(1) or AR (3).

- We then can inspect the plots at the within season lags, h=1,..,11. One perspective is that the ACF cuts off after 1 and the PACF tails off, and it indicates MA(1). Another perspective is that the ACF cuts off after 1 and the PACF cuts off after 4. In this situation, the book suggests to build a SARMA of orders p = 4 and q = 1. However, our professor points out that this is a bad reasoning. It is still however tempting for me to try this model since I lean to the side that the PACF cuts off after lag 4 instead of tails off (some subjective feeling goes in here). We right now don't know how to handle the situation where both ACF and PACF cuts off at a certain lag other than the approach mentioned in the book. So I tried this model. In diagnostic procedures, as you will see, it works better at some criterion. Together, I proposed two additional models and compare them to the one proposed by Travis.

- 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.

- While going through a bunch of models, the following model seems most appropriate as noted by everyone. sarima(econ[,2],0,2,1,1,1,0,12) with the following diagnostics: [image: Inline image 2] The adf test also suggests stationarity as follows: Also, I am working on other predictor variables to develop a preliminary regression model.

- I am trying to fit the regression model here and was wondering if I should take the stationary data for my fitting? Any help on this would be appreciated.

- I believe you would need to use the stationary non-seasonal unemployment rate as your response variable. I do not believe your predictor variables have to be stationary, but it would probably make sense to at least take the seasonality out of them.

- Checking both residual plots shows that there is a drastic drop that is likely indicative of the 2008 recession. Can we add weighting or something to fix this?

- In the political dataset I have an indicator variable for recession by month we could try using that.

- Regarding the expectations for presentation, the professor has not mentioned yet. However, in the last two lectures (14 and 15), he talked a lot applied examples about model building. I'd assume that our presentation would be something similar to what he talked in the two lectures. Basically, it's the model building process. How do we preprocess our data to obtain a stationary process (difference order 2 and difference order 1 in our case)? How do we identify the model (based on ACF and PACF)? What is the set of candidate models? How do you choose the best one (AIC, BIC, diagnostics)? I guess we might not need to present a regression model at this stage since he hasn't talked much about it. How do you guys think about this?

- "Best model," as far as I know, is pretty ambiguous right now. With what I have done before, I checked AIC and BIC (not really thinking about using R-squared for the time being). The model identification from P/ACF is outlined in the text by checking out the tail behavior to see if it decays asymptotically or cuts off. We should be checking inside the band for "cutoff" behavior.

- Sure, it's always hard to call a model "Best". I think in presentations, we may present several potential candidate models, and compare them from several perspectives. Hopefully, one model will gain relatively more evidence.

- I would like to propose an additional model. I have gone through the same exercise as @trlilley12 and @bopangpsy only I used the seasonally adjusted unemployment rate. It looks like the performance is definitely comparable to the seasonal models. I used sarima for the nice diagnostic plot it creates, but I left the seasonal parameters out.

I get an AICc $= -2.617631$ and BIC $= -3.598962$

```
mdl4 = sarima(xdata = unem, p = 1, d = 2, q = 1)

Call:
stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
Q), period = S), include.mean = !no.constant, optim.control = list(trace = trc,
REPORT = 1, reltol = tol))

Coefficients:
ar1      ma1
-0.2021  -0.8078
s.e.   0.0688   0.0433

sigma^2 estimated as 0.02626:  log likelihood = 109.15,  aic = -212.3

$AIC
[1] -2.625197
```

5

```
$AICc
[1] -2.617631

$BIC
[1] -3.598962
```

Cool, Joseph! This model is simple and performs pretty well in terms of both fitting indices and diagnostics.
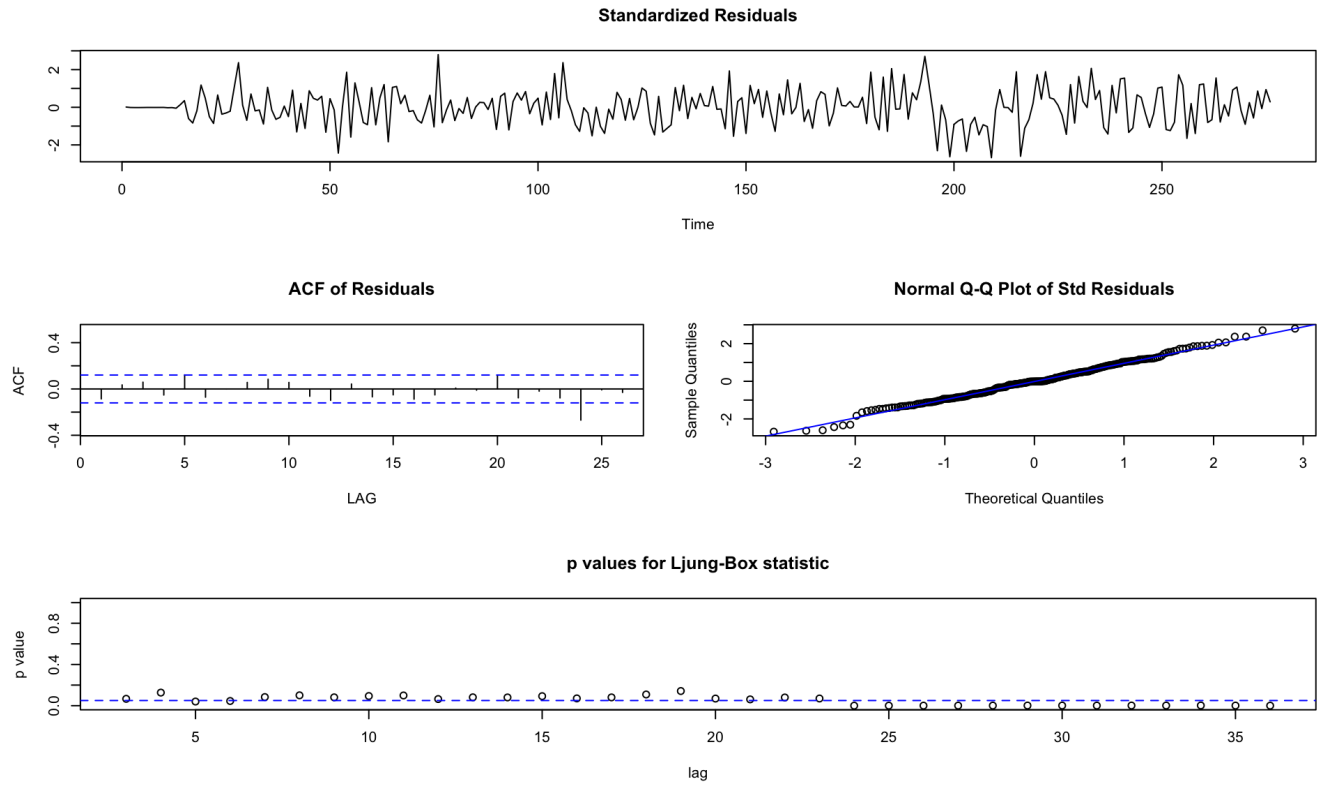
2.1 Overview
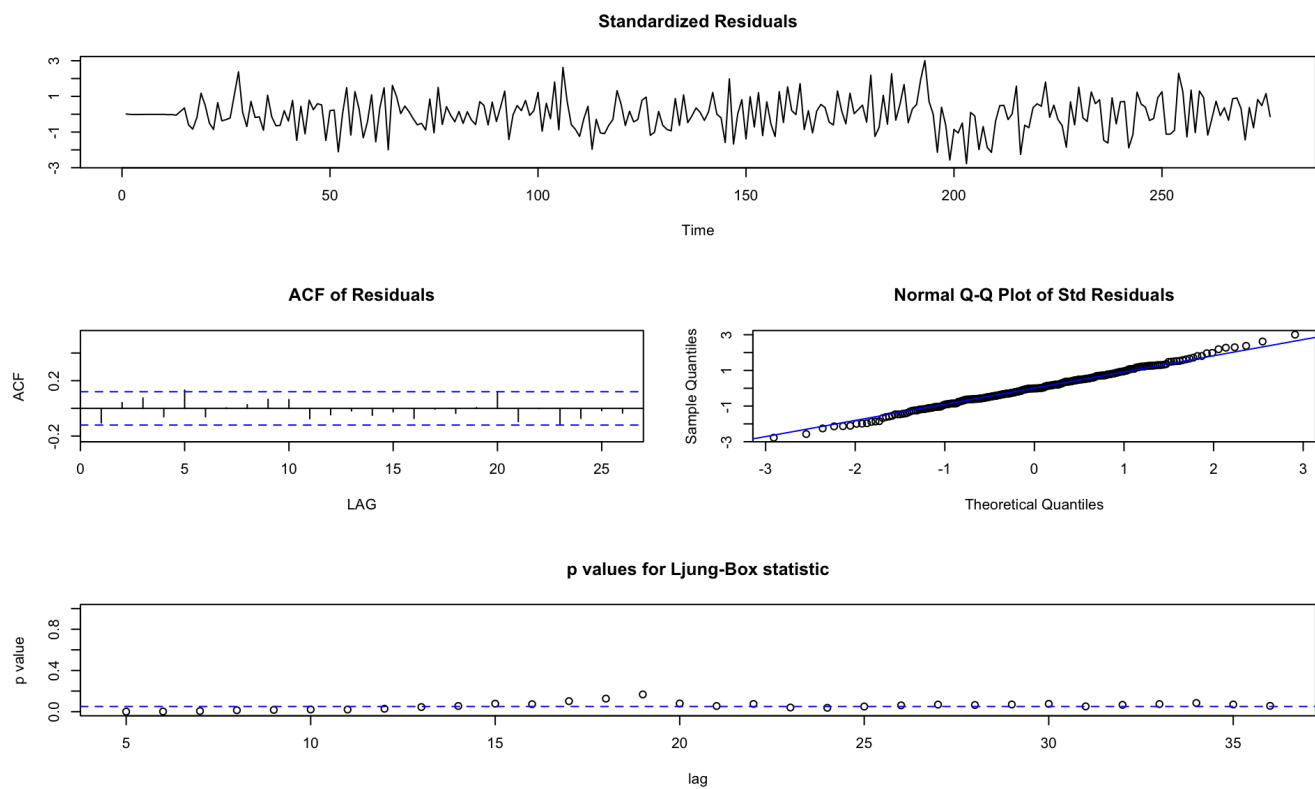
Models Considered

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$(0, 2, 1) \times (1, 1, 0)_{12}$

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

Standardized Residuals

ACF of Residuals

Normal Q-Q Plot of Std Residuals

p values for Ljung-Box statistic
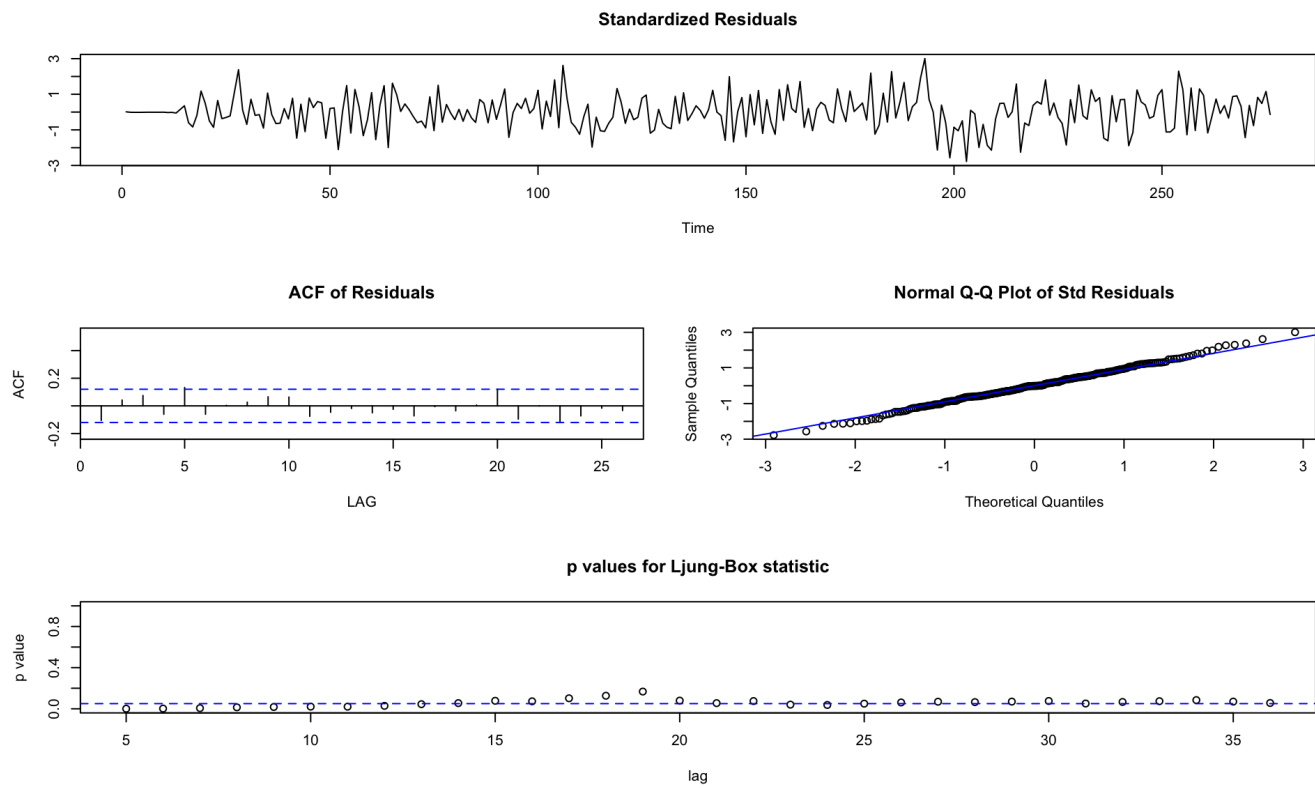
2.3   Seasonally Adjusted Models

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

8
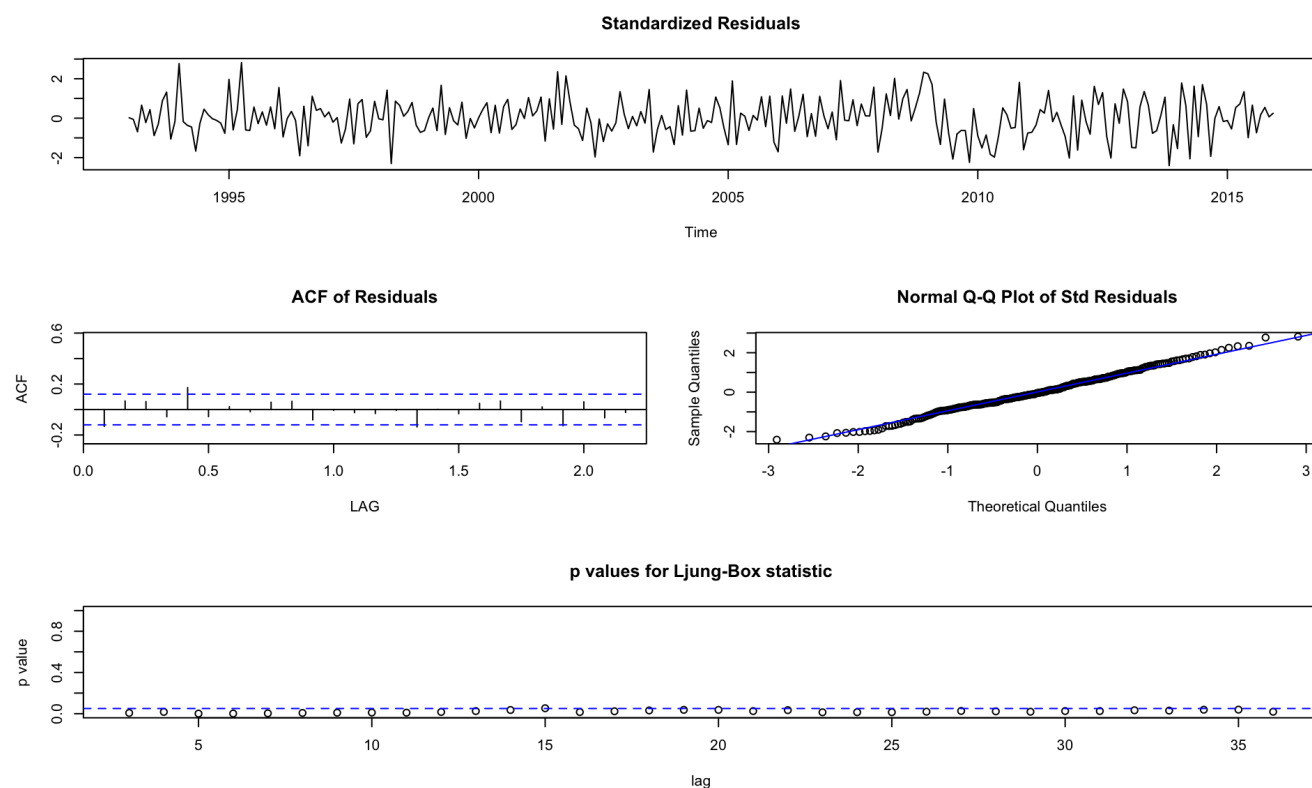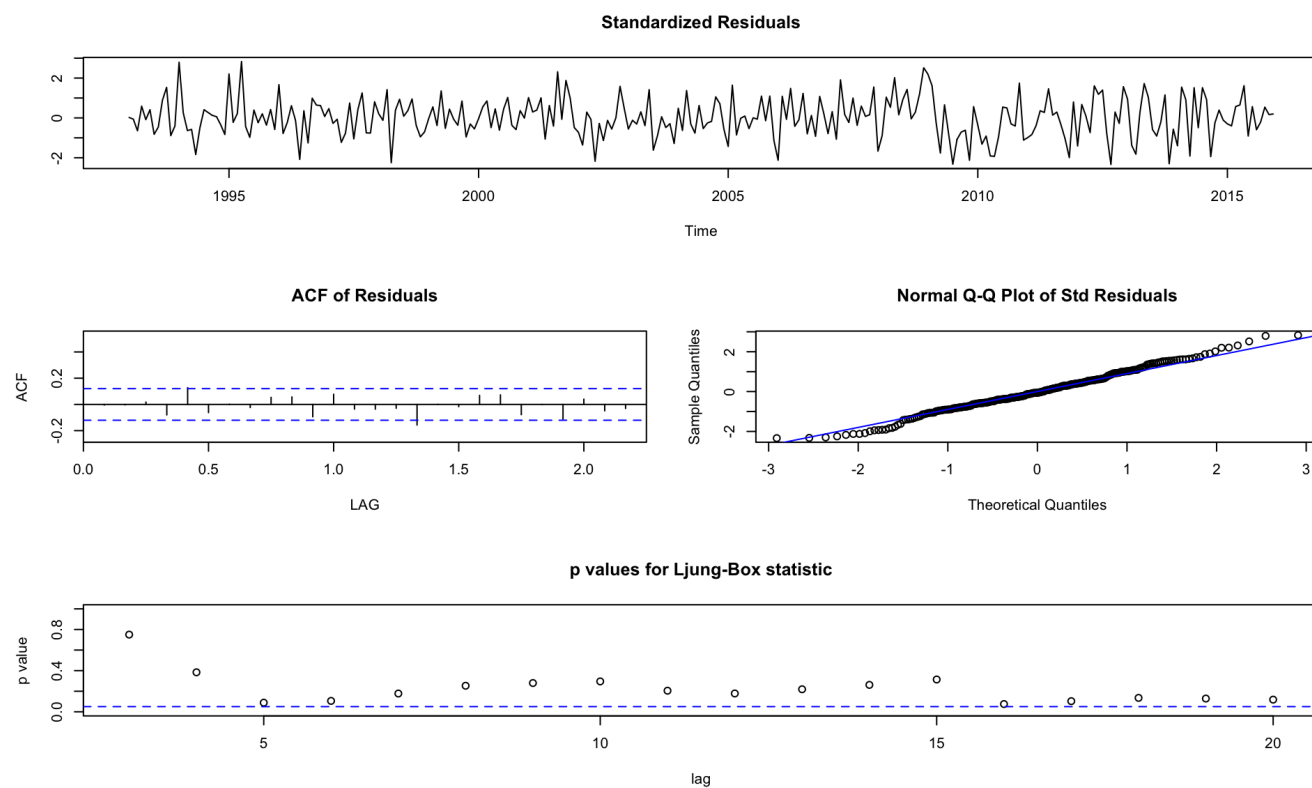
Model 6: SARIMA$(0, 2, 1) \times (1, 0, 0)_{12}$ with Regressors

Standardized Residuals

ACF of Residuals

Normal Q-Q Plot of Std Residuals

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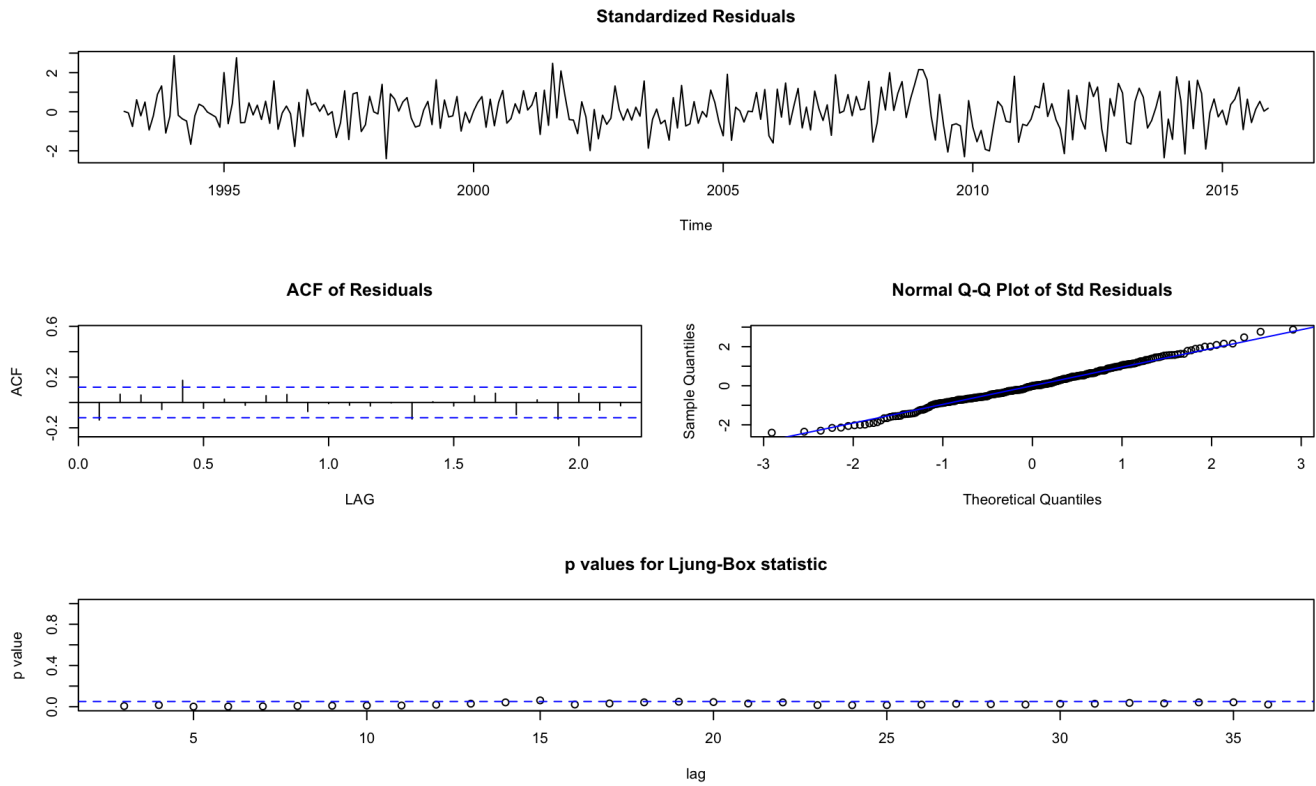
Model 7: ARIMA$(1, 2, 1)$ with Regressors

Standardized Residuals

ACF of Residuals

Normal Q-Q Plot of Std Residuals

p values for Ljung-Box statistic