James Pilarski
Time Series Analysis Student Project
Spring 2012 Exam Semester

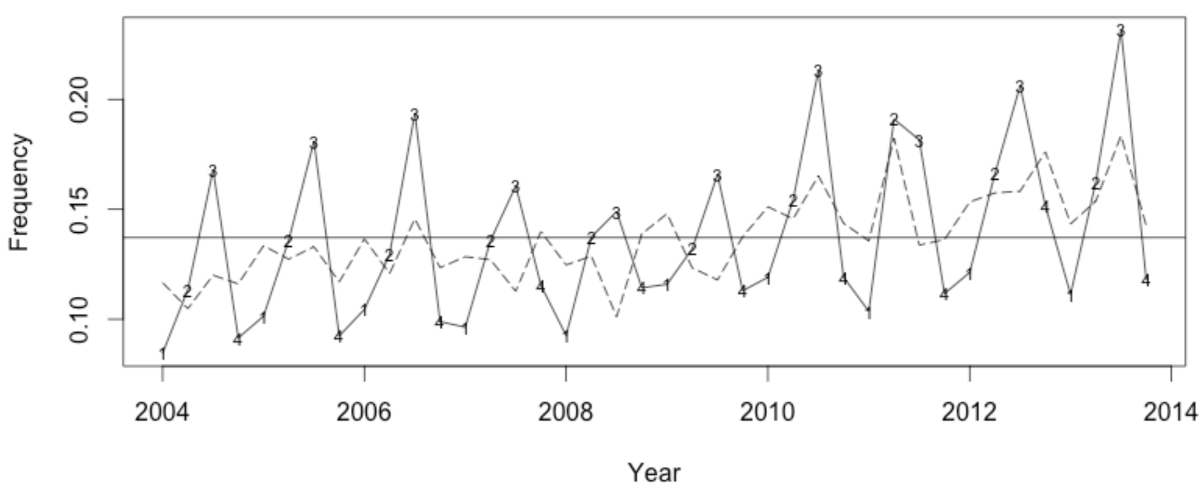
Introduction

Actuaries are commonly responsible for estimating loss trend when deriving projected loss estimates for a specified future period. Trends are typically estimated by fitting an exponential curve to the data. This method offers convenience but it is not terribly accurate when the underlying loss trend is seasonal, as is the case with the data examined in the time series analysis outlined below. This project will focus on one component of the loss trend, claim frequency, and compare how well a model fits unadjusted and seasonally adjusted times series.

Data Evaluation

The data to be used in this project include 40 quarterly evaluations of reported claims from January 2004 to December 2013. The claim counts at each evaluation are divided by the number of exposures earned in that period to yield claim frequencies free of distortions caused by changes in the amount of business written from one quarter to the next. Figure 1 contains a time series plot of claim frequencies by quarter. The solid line represents unadjusted observations while seasonally adjusted frequencies follow the dashed line.

Figure 1: Quarterly Claim Frequency

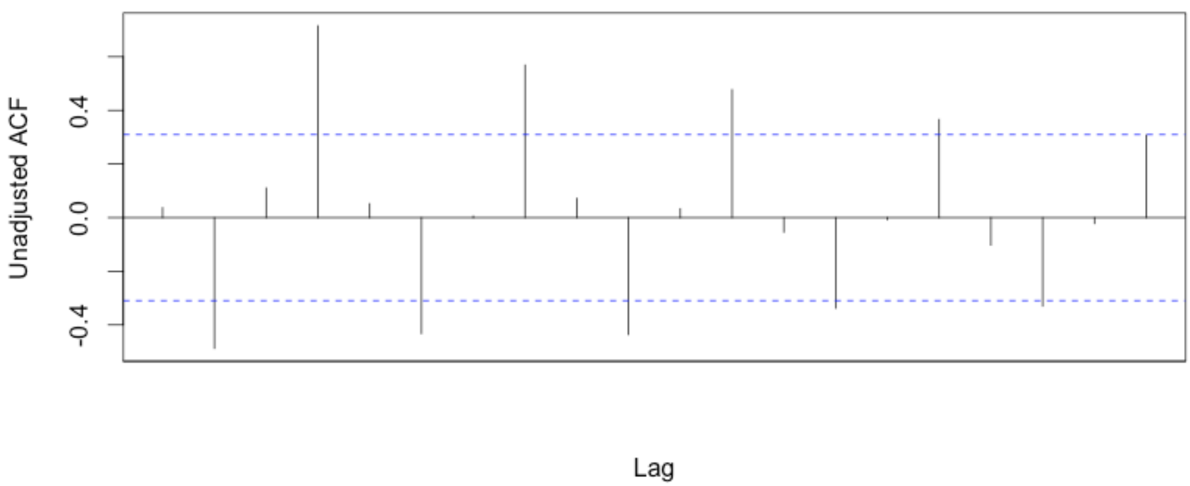


Several observations can be made using this plot with the most obvious being a strong seasonal pattern in the unadjusted claim frequency. The seasonally adjusted data exhibit no such pattern. However, both time series possess upward trends relative to the horizontal line representing the average claim frequency over the 10 year period. Trend of this nature is indicative of nonstationary time series and will need to be removed prior to specifying the appropriate time series model.

Model Specification

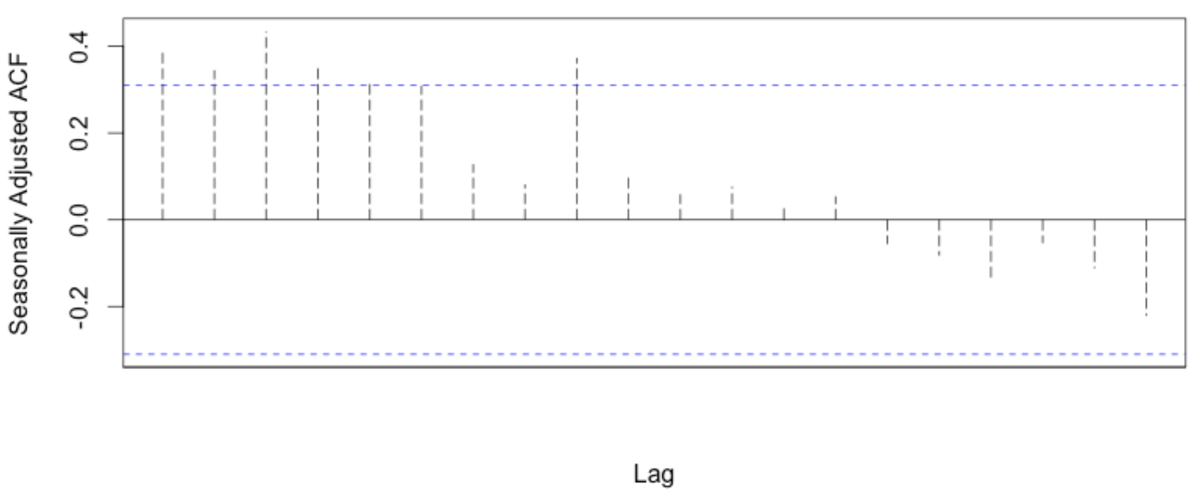
When deciding which time series models to consider, it is necessary to examine the sample autocorrelation and sample partial autocorrelation functions, abbreviated ACF and PACF, respectively. A plot of the ACF for unadjusted claim frequencies is shown in Figure 2 and clearly exhibits an oscillating seasonal pattern with a period of 4 quarters. This graph also provides additional evidence that the time series is nonstationary since the autocorrelations are decreasing slowly as the lag increases.

Figure 2: Sample ACF of Claim Frequency



The autocorrelation pattern in Figure 3 lacks oscillation as it was constructed using seasonally adjusted claim frequencies. The plot also supports the conclusion that trend is present as it too fails to decrease rapidly.

Figure 3: Sample ACF of Claim Frequency



A review of the PACF plots for unadjusted and seasonally adjusted claim frequencies can be found in Figures 4 and 5, respectively. In both cases, there are two statistically significant values that would indicate one should incorporate between 2 and 9 autoregressive parameters in the time series models. This seems unreasonable and is likely to change once they are made stationary by removing the trend.

Figure 4: Sample PACF of Claim Frequency

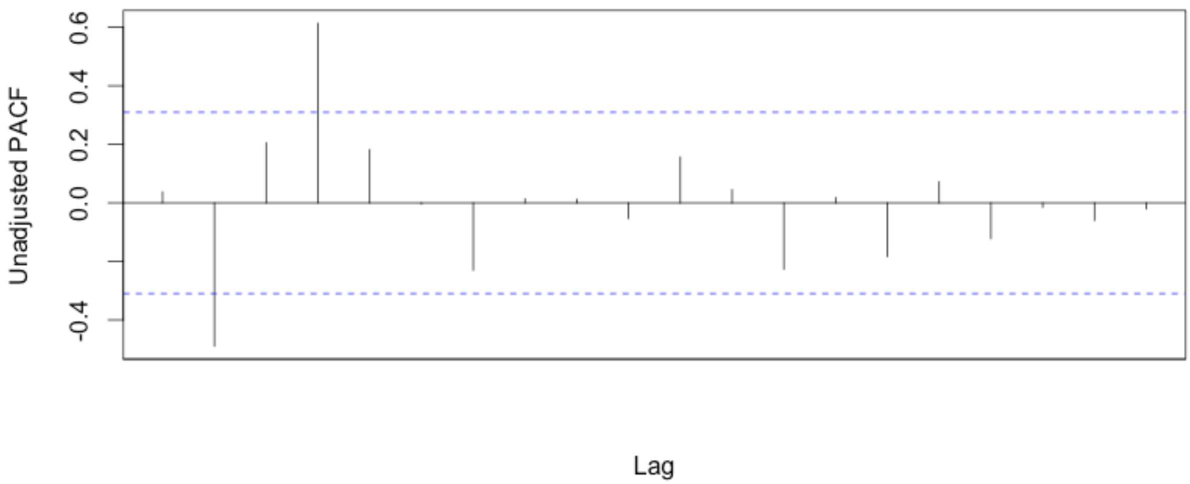
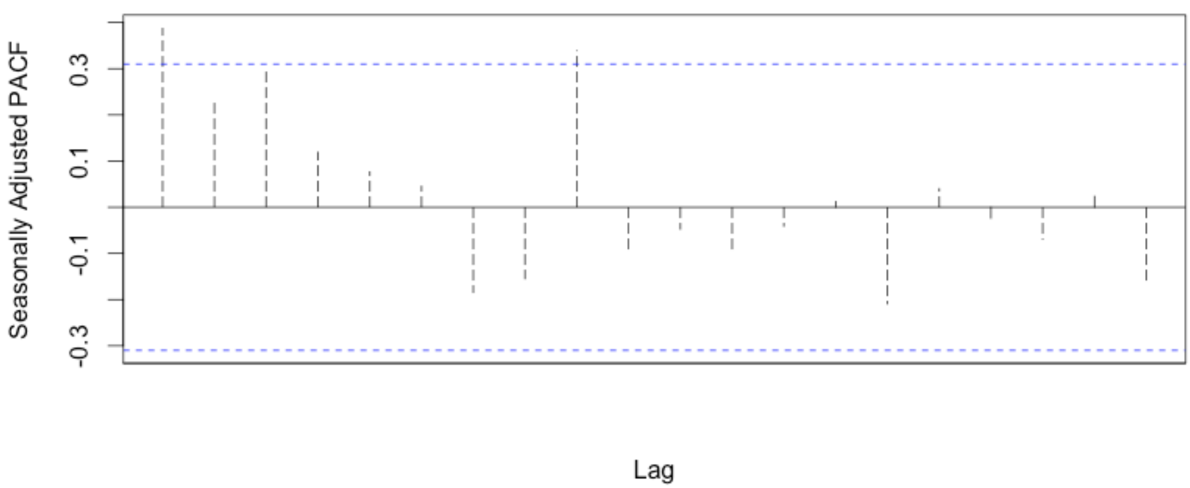
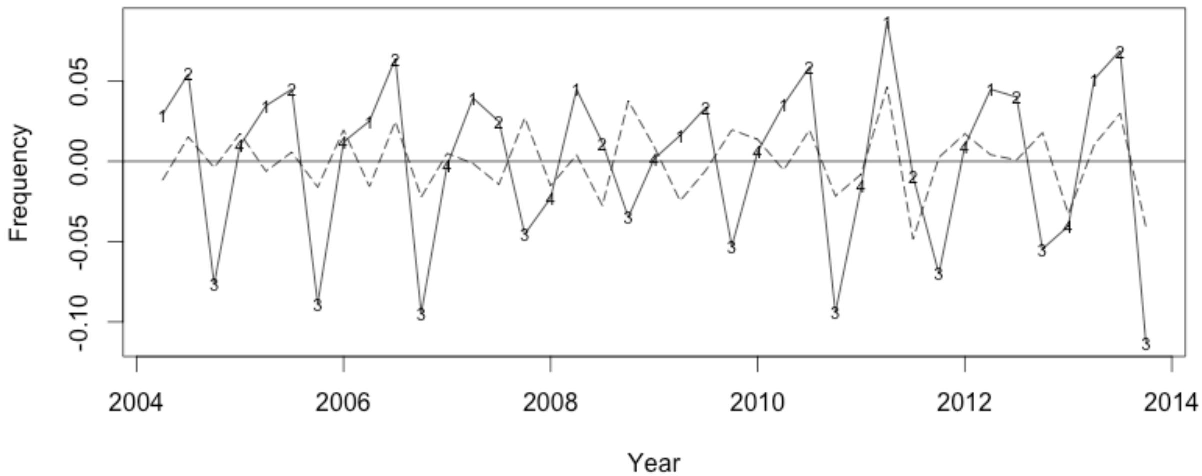


Figure 5: Sample PACF of Claim Frequency



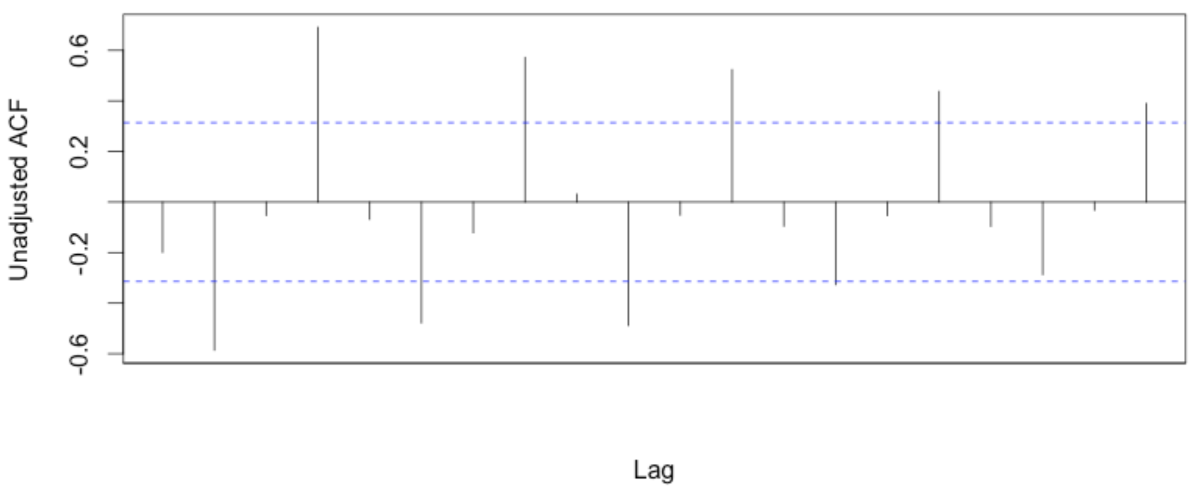
It is possible to remove the trends from a time series through differencing. Both the unadjusted and seasonally adjusted data are nonstationary and were modified using the first difference with a lag of 1. These results are displayed in Figure 6 where both series are no longer exhibiting upward trends. A seasonal pattern appears to persist in the unadjusted observations.

Figure 6: First Difference of Quarterly Claim Frequency at Lag 1



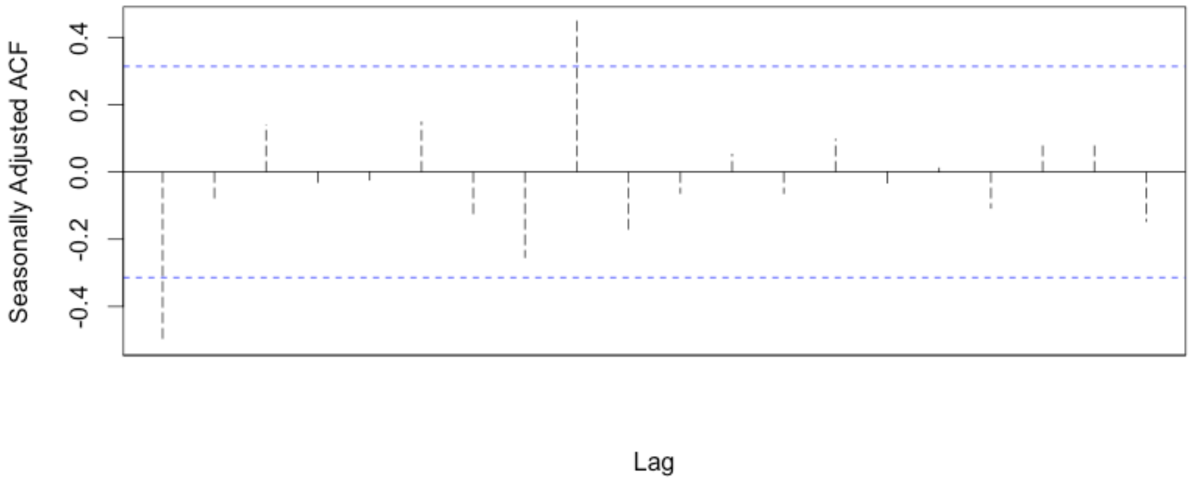
Once again it will be useful to examine the ACF plots of these time series. Figure 7 displays the unadjusted data detrended through differencing. The seasonal pattern remains and will require additional attention.

Figure 7: Sample ACF of the First Difference at Lag 1 of Frequency



The seasonally adjusted ACF plot in Figure 8 shows much more encouraging results since the autocorrelation decreases fairly rapidly after the first lag. There is an obvious exception at lag 9 which could indicate a seasonal MA(1) parameter is needed or it may simply be a random fluctuation.

Figure 8: Sample ACF of the First Difference at Lag 1 of Frequency



The corresponding graphs of the PACF are shown below in Figures 9 and 10 where they still exhibit some seasonality. More specifically, the appropriate models for the seasonally adjusted time series are AR(1) or AR(2) due to the large partial autocorrelations at lags 1 and 2. A spike in the PACF plot at lag 8 could suggest a seasonal AR(1) should added to the model if this is not simply the result of randomness.

Figure 9: Sample PACF of the First Difference at Lag 1 of Frequency

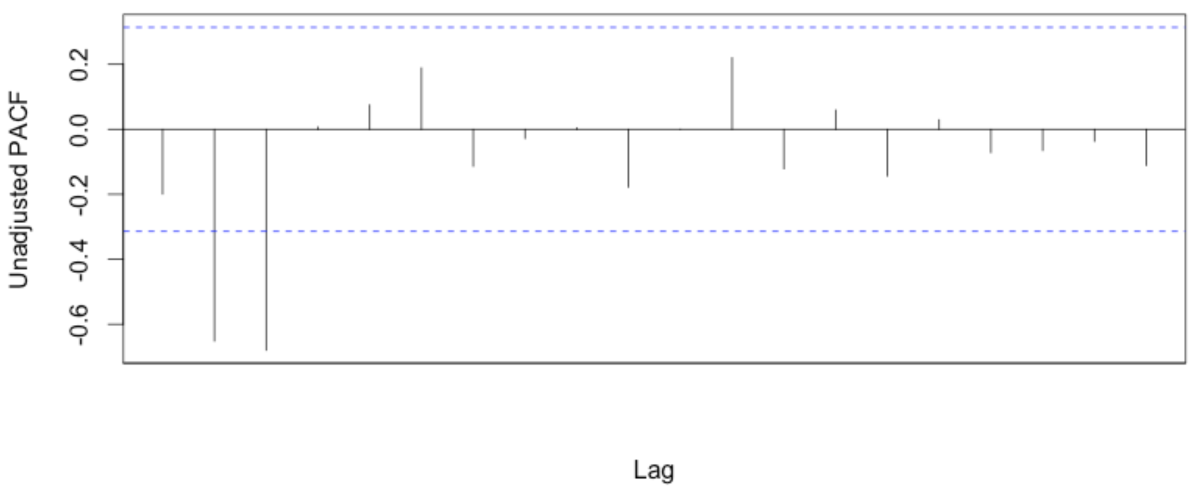
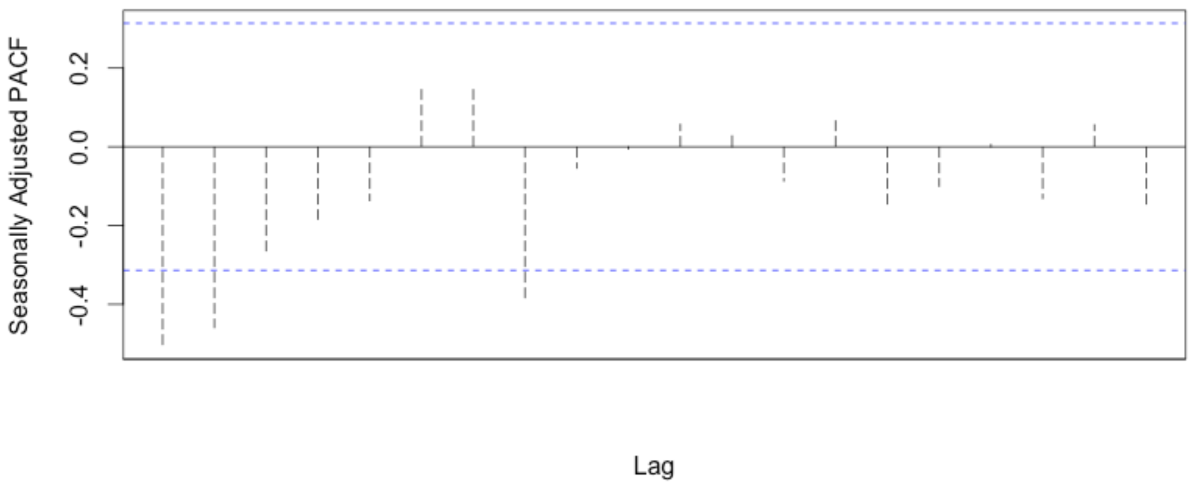
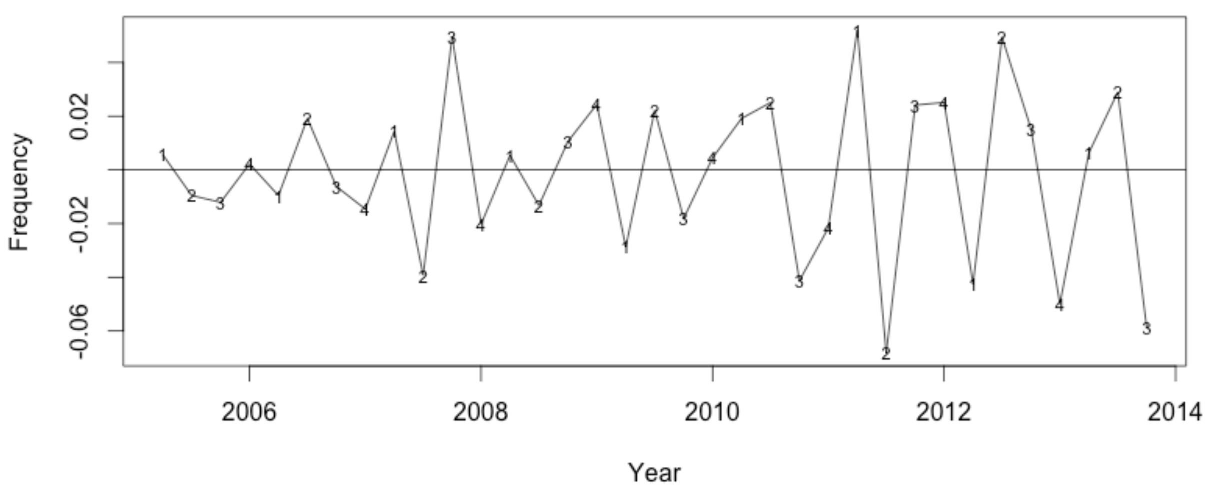


Figure 10: Sample PACF of the First Difference at Lag 1 of Frequency



Returning to unadjusted observations, the quarterly pattern can be removed from the time series through seasonal differencing with a lag of 4. Figure 11 shows the effect of this modification on the data.

Figure 11: First Difference of Quarterly Claim Frequency at Lags 1 and 4



The ACF plot in Figure 12 now contains a statistically significant autocorrelation at lags 1, 3, and 9 with nearly significant values at lags 4 and 8. This pattern seems to indicate the time series model should include both nonseasonal and seasonal MA(1) parameters.

Figure 12: Sample ACF of the First Difference at Lags 1 and 4 of Frequency

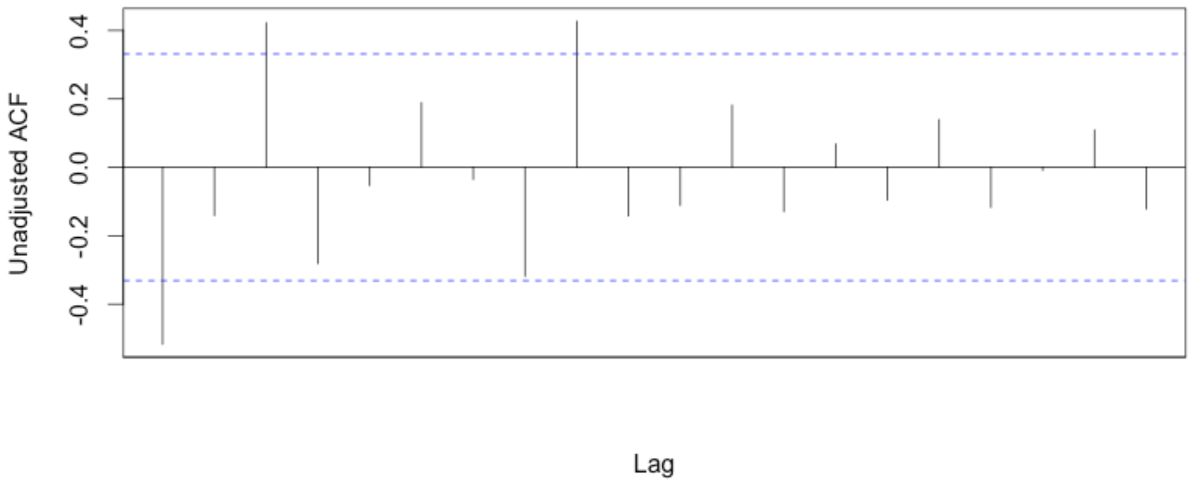
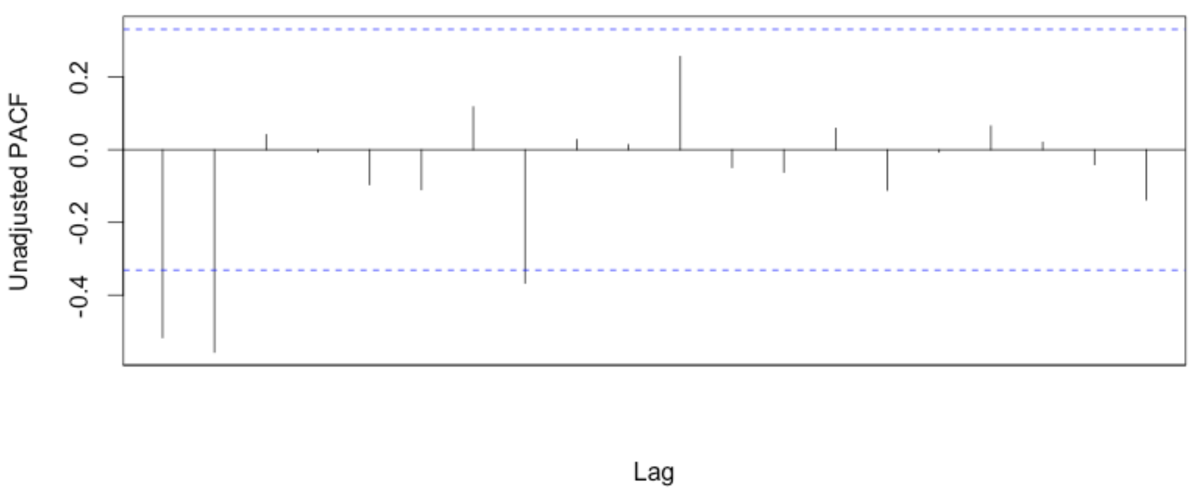


Figure 13 displays a PACF plot that seems to suggest the a nonseasonal AR(2) parameter should be added to the model as well as a seasonal AR(1) parameter.

Figure 13: Sample PACF of the First Difference at Lags 1 and 4 of Frequency



When the above observations are pulled together, the unadjusted time series will be fit to an ARIMA(2,1,1)x$(1,1,1)_4$ model while an ARIMA(2,1,1)x$(1,0,1)_4$ model will be specified for the seasonally adjusted data. It is interesting to note the only difference between these models is the lack of seasonal differencing in the one applied to the seasonally adjusted data.

Model Fitting

The models selected above were fit using maximum likelihood estimation. Table 1 contains the fitted parameters for the unadjusted time series while Table 2 displays those for the seasonally adjusted data. The initial fits revealed the seasonal AR(1) parameter is not statistically different from zero. Therefore, this parameter was removed from the models in order to avoid overfitting.

Table 1: Unadjusted Time Series

Series: freq
ARIMA(2,1,1)(1,1,1)[4]

Coefficients:

	ar1	ar2	ma1	sar1	sma1
	-0.3896	-0.2821	-0.6856	0.1513	-0.9999
s.e.	0.2347	0.2212	0.2471	0.1919	0.2475

sigma^2 estimated as 0.0002221
log likelihood=92.58
AIC=-173.16 AICc=-170.16 BIC=-163.83

Table 2: Seasonally Adjusted Time Series

Series: freq.adj
ARIMA(2,1,1)(1,0,1)[4]

Coefficients:

	ar1	ar2	ma1	sar1	sma1
	-0.4476	-0.3236	-0.5046	-0.3595	0.5077
s.e.	0.2548	0.2121	0.2381	0.7492	0.6903

sigma^2 estimated as 0.0002182: log likelihood=108.4
AIC=-204.8 AICc=-202.18 BIC=-194.82

Tables 3 and 4 below show the results of refitting the models. Removing the seasonal AR(1) parameter decreased the AIC and BIC for both models possibly indicating better fits. It is worth noting the seasonal MA(1) parameter in the seasonally adjusted model is also not statistically different from zero and its removal may further improve the fit.

Table 3: Unadjusted Time Series

Series: freq
ARIMA(2,1,1)(0,1,1)[4]

Coefficients:

	ar1	ar2	ma1	sma1
	-0.3742	-0.2796	-0.6792	-0.9999
s.e.	0.2133	0.2081	0.2048	0.4399

sigma^2 estimated as 0.0002194
log likelihood=92.26
AIC=-174.52 AICc=-172.45 BIC=-166.74

Table 4: Seasonally Adjusted Time Series

Series: freq.adj
ARIMA(2,1,1)(0,0,1)[4]

Coefficients:

	ar1	ar2	ma1	sma1
	-0.4703	-0.3343	-0.4979	0.1561
s.e.	0.2519	0.2128	0.2410	0.1947

sigma^2 estimated as 0.0002197
log likelihood=108.28
AIC=-206.55 AICc=-204.73 BIC=-198.23

As expected, the AIC and BIC decreased and the remaining parameters did not change much relative to their standard errors. These statistics can be found in Table 5 below. This result is not surprising in light of the seasonal adjustment made to the underlying data. One may conclude the significant values identified at the later lags in the ACF and PACF plots above are likely the result of random fluctuation. Now that both models have been fit, it is necessary to assess the quality of that fit using a series of diagnostics.

Model Diagnostics

Many diagnostic tests of time series models examine the residuals that result from the difference between actual and fitted observations. The residuals of well fit models should mirror a normally distributed white noise process. A plot of the residuals should therefore appear as a patternless scatter of points. The residuals from the unadjusted frequencies can be observed in Figure 14 where they appear to have an upward trend. In Figure 15, one can see the model fit to the seasonally adjusted data produces residuals that appear much more similar to white noise.

Table 5: Seasonally Adjusted Time Series

```
Series: freq.adj
ARIMA(2,1,1)

Coefficients:
         ar1      ar2      ma1
     -0.4207  -0.3032  -0.5171
s.e.  0.2203   0.1948   0.1945
sigma^2 estimated as 0.0002239
log likelihood=107.98
AIC=-207.96   AICc=-206.78   BIC=-201.31
```

Figure 14: Residual Plot of ARIMA(2,1,1)x(0,1,1)[4]

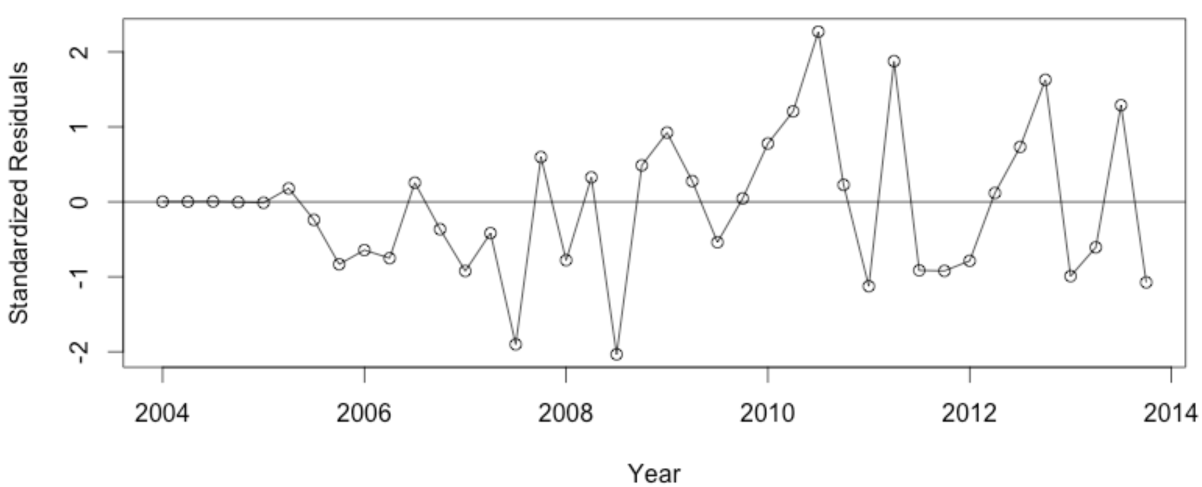
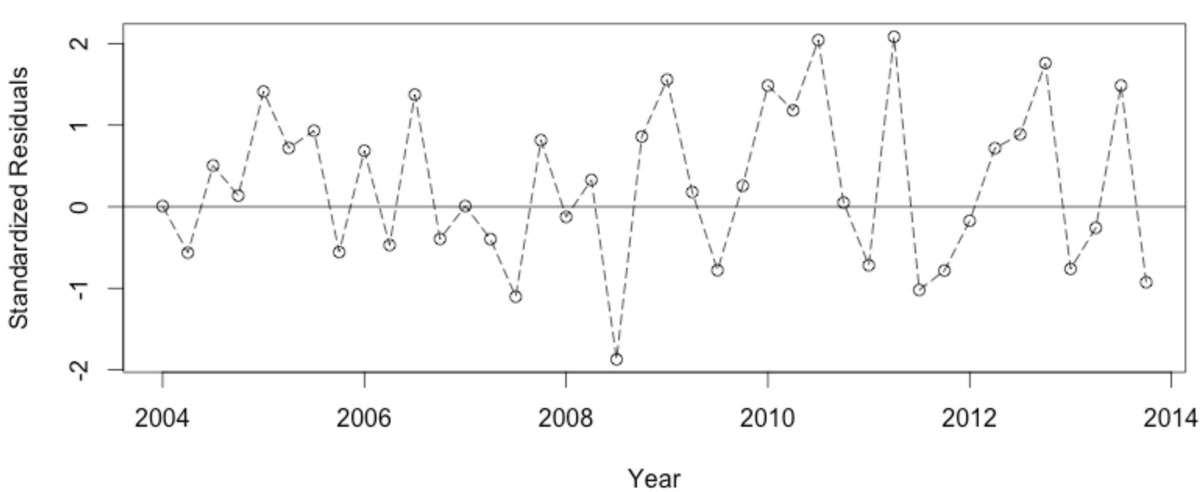


Figure 15: Residual Plot of ARIMA(2,1,1)



The extent to which the residuals follow a normal distribution can be evaluated by viewing a histogram of the residuals and quantile-quantile plots. Figures 16 and 17 show histograms that are similar to the shape of a normal probability distribution function but perhaps not enough.

Figure 16: Histogram of Residuals

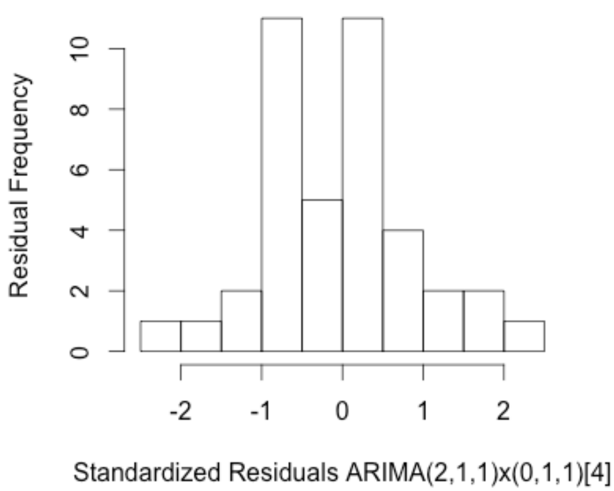
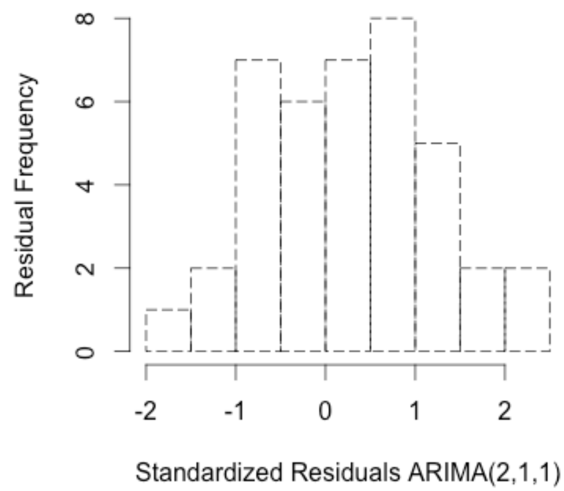


Figure 17: Histogram of Residuals



The quantile-quantile residual plots from the two models are shown in Figures 18 and 19. Neither model produces residuals that follow a normal distribution perfectly since plotted points do not closely follow the diagonal lines through the graphs. That said, the seasonally adjusted data produce residuals that are slightly more normal.

Figure 18: Normal Q-Q of Residuals

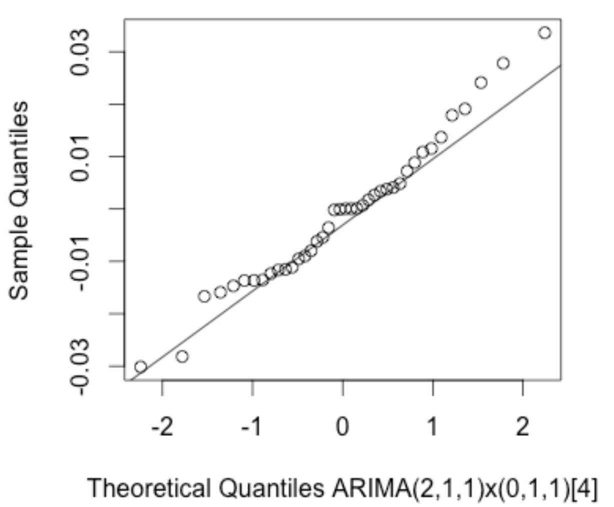
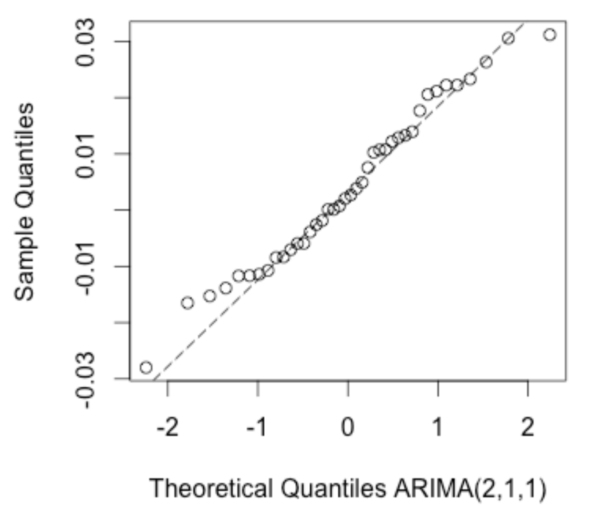
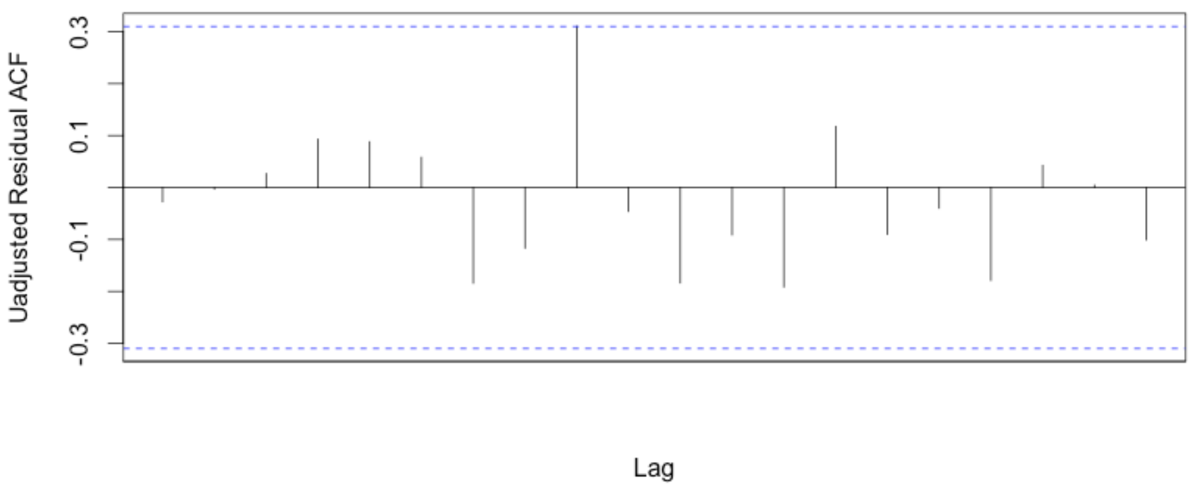


Figure 19: Normal Q-Q of Residuals



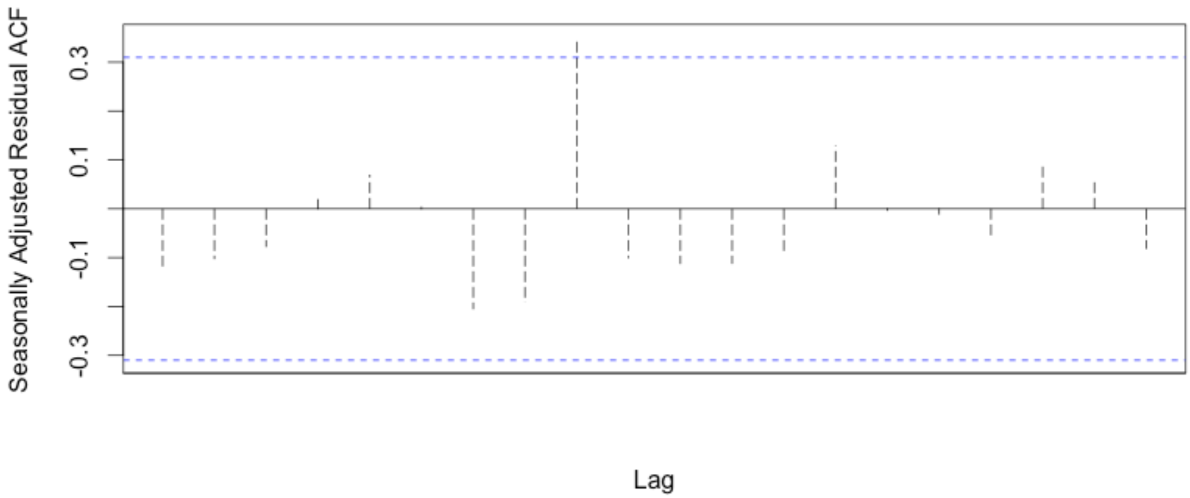
An additional diagnostic of model fit examines the independence of the residuals. Calculating the autocorrelation of residuals and plotting the results can reveal dependencies in the error terms. Figure 20 contains the ACF plot of residuals from the unadjusted time series. With the exception of lag 9, there are no statistically significant correlations among the error terms for this model.

Figure 20: Sample Autocorrelation of Residuals from ARIMA(2,1,1)x(0,1,1)[4]



The residuals produced by the model fit to the seasonally adjusted claim frequencies also show a significant autocorrelation at lag 9 as one can see in Figure 21. Again, this may be the result of random fluctuation.

Figure 21: Sample Autocorrelation of Residuals from ARIMA(2,1,1)



Further examination of the independence of residuals can be performed by viewing PACF plots. These can be seen in Figures 22 and 23 which do not contain any statistically significant values but lag 9 is still the largest value in both plots.

Figure 22: Sample Autocorrelation of Residuals from ARIMA(2,1,1)x(0,1,1)[4]

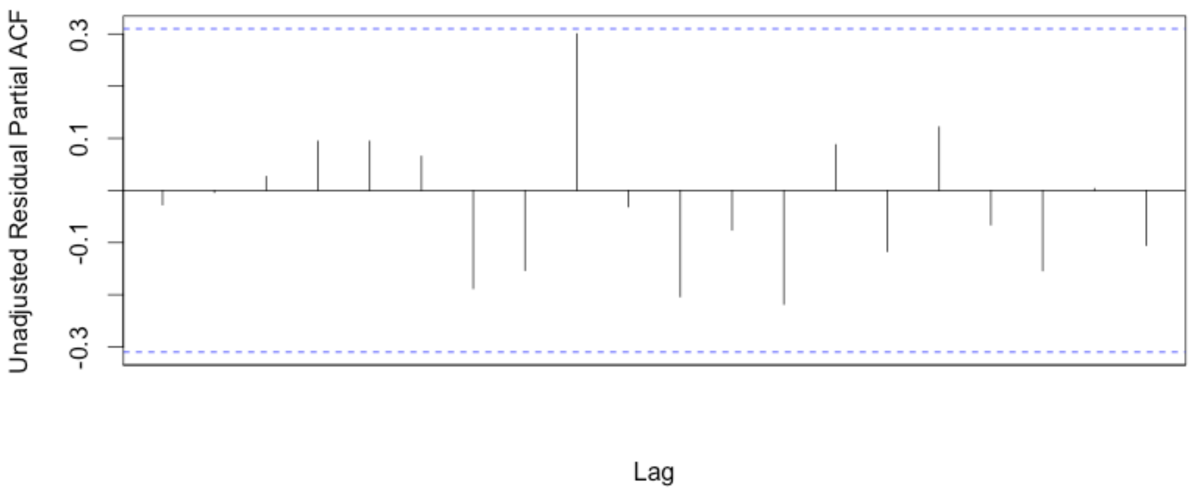
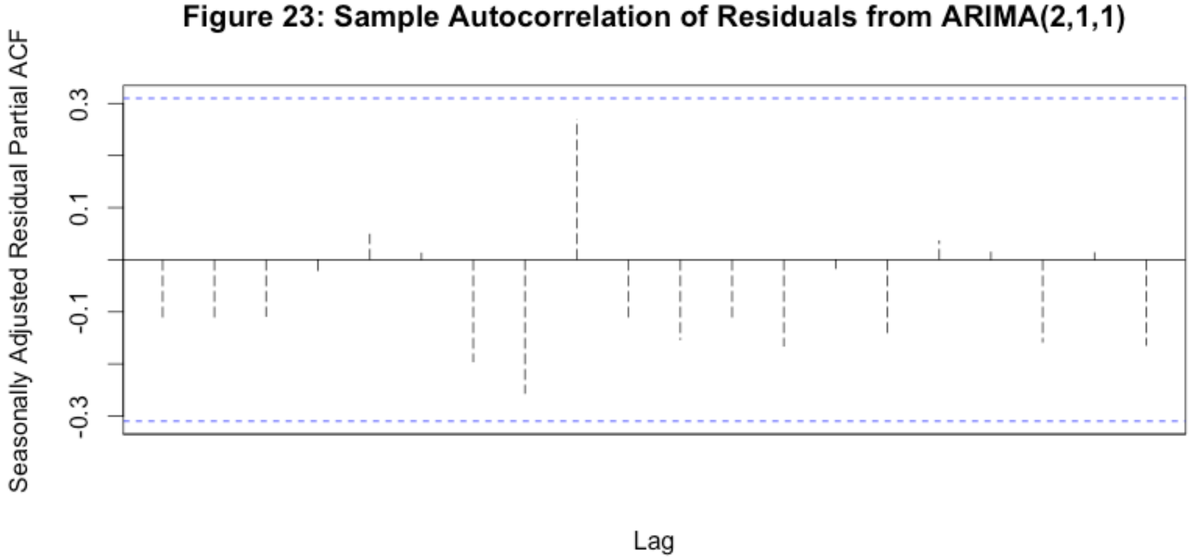


Figure 23: Sample Autocorrelation of Residuals from ARIMA(2,1,1)



A final assessment of correlation among residuals can be made using the Box-Ljung test which examines the magnitude of residual correlations for more than one lag at a time. Tables 6 and 7 show the calculated statistics and the corresponding p-values for each set of observations. Based on these figures, the null hypothesis that states the residuals are correlated would be rejected a significance of 20% but not 15% when one tests the unadjusted observations, not a terribly convincing case. The null hypothesis for the seasonally adjusted time series would be rejected at 10% but not 5%, a much better result.

Table 6: Unadjusted Time Series

Box-Ljung test
data: residuals(fit02)
X-squared = 8.8013, df = 6,
p-value = 0.1851

Table 7: Seasonally Adjusted Time Series

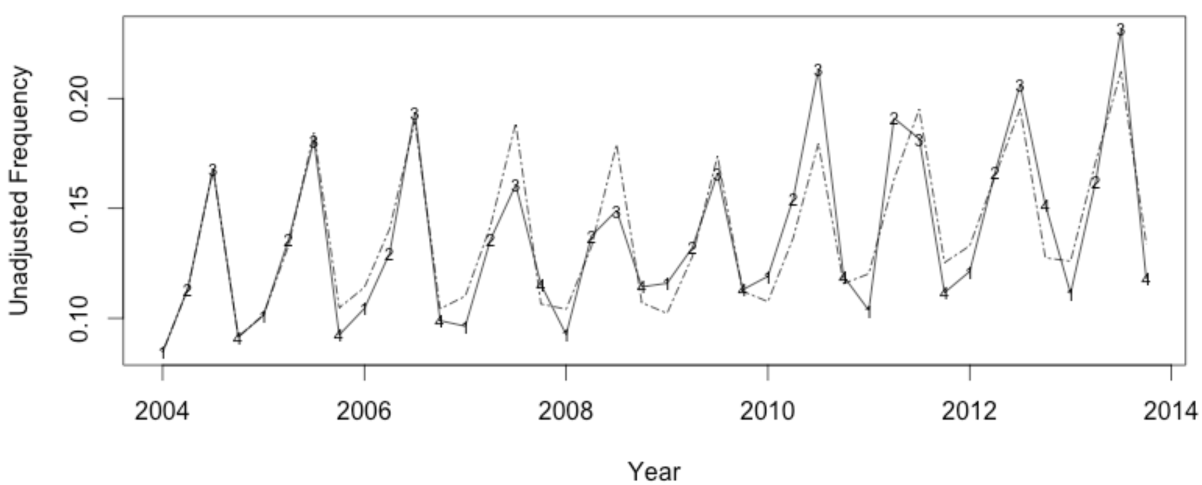
Box-Ljung test
data: residuals(fit03.adj)
X-squared = 12.9624, df = 7,
p-value = 0.07303

The analysis outlined above would lead one to conclude that the fit of the ARIMA(2,1,1) with seasonally adjusted data is better than that of ARIMA(2,1,1)x$(0,1,1)_4$ model for the unadjusted time series. Additional visual inspection can help verify this conclusion.

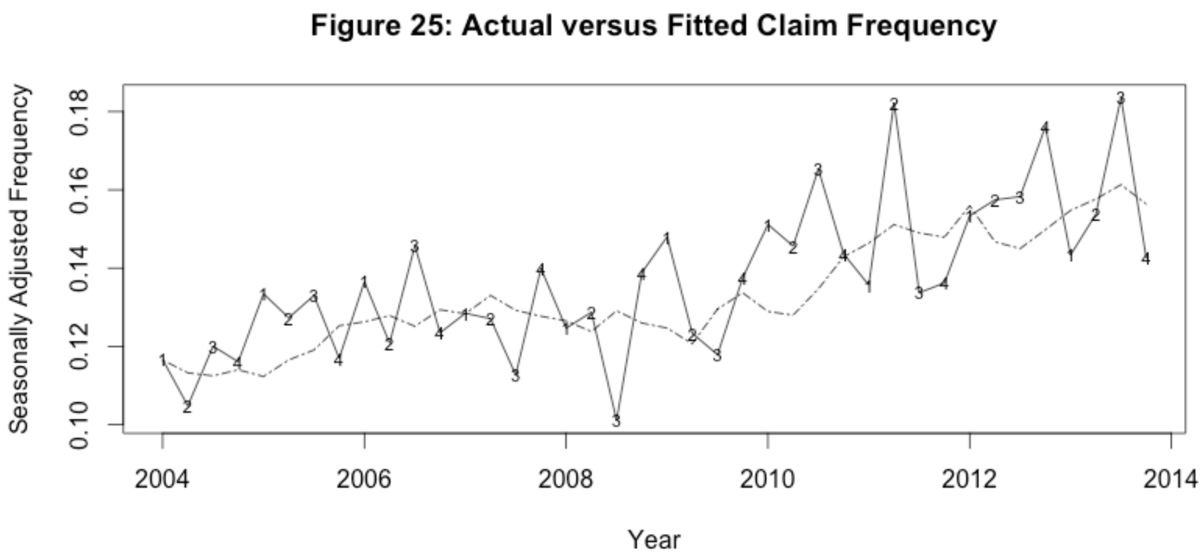
<u>Actual versus Fitted</u>

Now that the time series models have been specified, fitted, and examined for quality, one can plot the actual versus fitted observations. Figure 24 shows this plot for the unadjusted claim frequency and the fit appears to be rather good.

Figure 24: Actual versus Fitted Claim Frequency



The same cannot be said for the seasonally adjusted fit in Figure 25. This is an unexpected result given the tests of fit performed up to this point in the analysis.



Conclusion

The visually poor fit of the model for seasonally adjusted claim frequency could be the result of many factors. One possibility is the statistically significant residual autocorrelations at lag 9 represent a seasonal factor that was not accounted for properly. The seasonal pattern in these data seems to shift slightly at certain times and a smoothing adjustment may improve the fit. Whatever the case, it is clear that fitting seasonal models is rather difficult and must require techniques that are beyond to scope of the lessons covered in this course.