NEAS Time Series
Dorothy Fong
Summer 2011 Session
dorothyfong@gmail.com

Student Project:
Modeling New Brunswick Automobile Industry Collision Claim Cost

1) Introduction

The goal of this project is to analyze and model semi-annual average claim cost for the collision coverage for the automobile industry in New Brunswick using the techniques introduced in the NEAS Time Series Course.

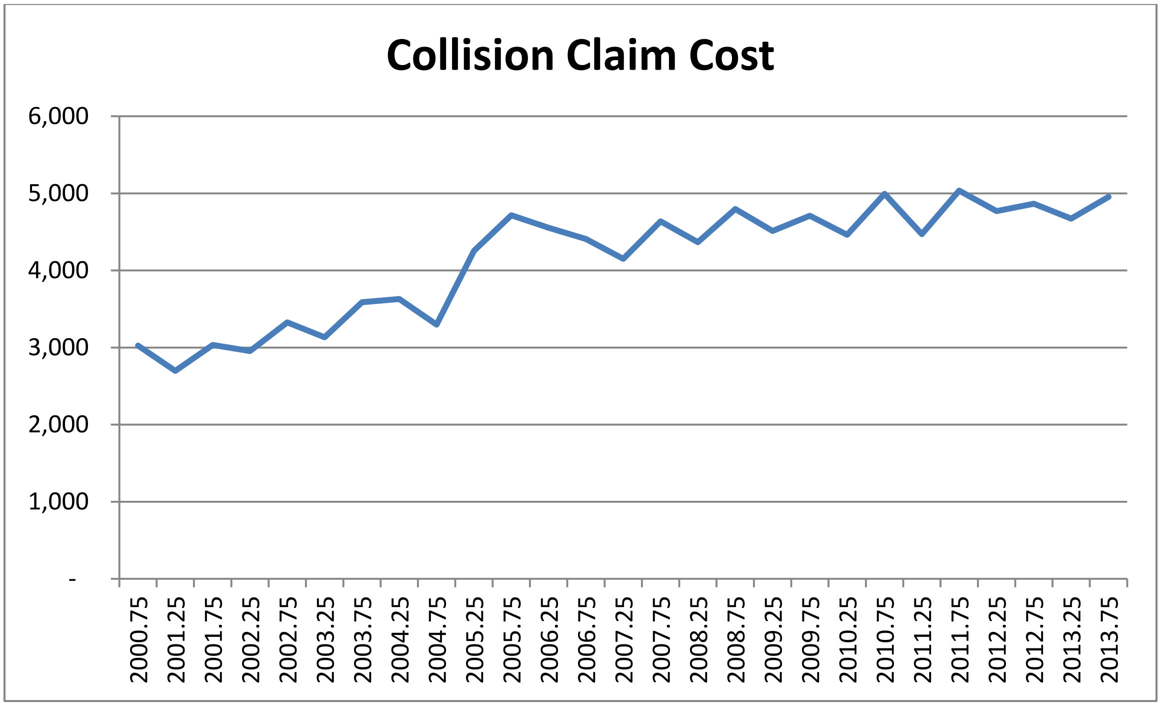
2) Data

Data used in this project is New Brunswick Automobile Industry wide Private Passenger Automobile claims data for Collision coverage. Data is as of December 31, 2013, and data points are presented on a semi-annual basis. The data included in the analysis ranges from December 2000 to December 2013. Since this is industry data for the province of New Brunswick, the data size is large and possesses full credibility. Data is examined for reasonableness and compared with Industry exhibits to ensure accuracy.

Time index 0.25 represents June data point and 0.75 represents December data point.
(E.g. 2001.25 represents June 2001, and 2001.75 represents December 2001.)

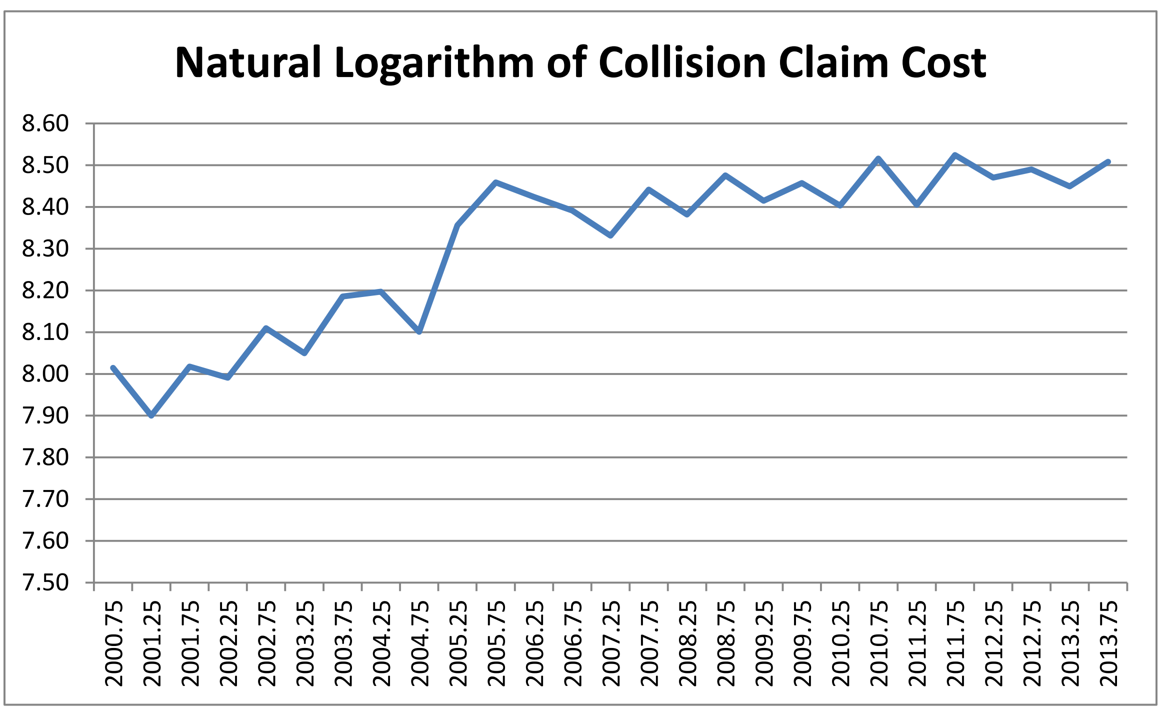
The data shows an upward trend over the period analyzed, reflecting inflation and possibly change in deductible levels throughout the years.

Collision claim costs experienced a bigger increase in 2005 as shown in the graph below.



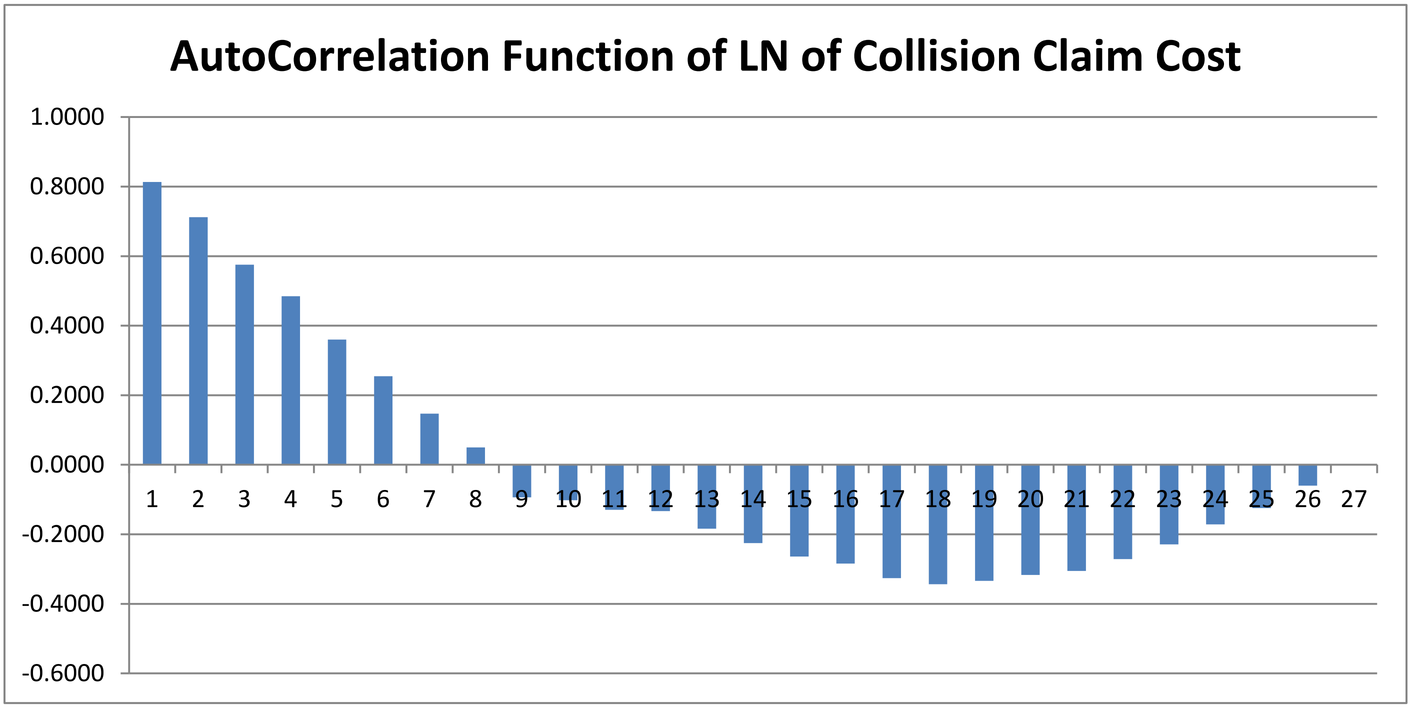
The natural logarithm of the collision claim cost graph is then examined, as lognormal distributions are typically used to model claim sizes.

The graph of the natural logarithm shows the same trend we observe in the original data.



3) Model Specification

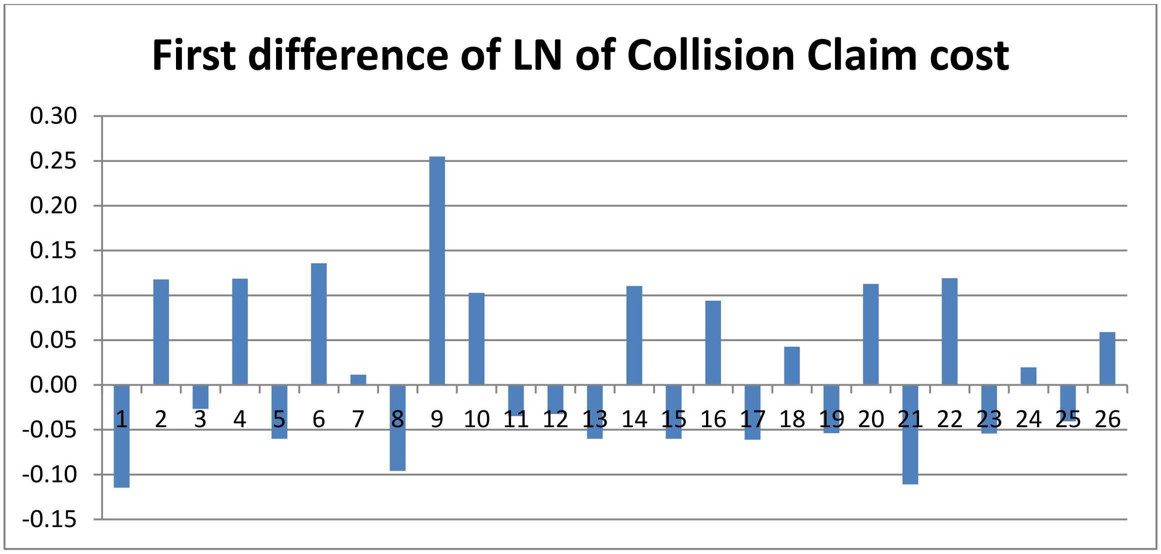
Autocorrelation function of the natural logarithm of the data is then reviewed:



Seasonality does not present in the data as there are no evident peaks in our graph. ACF tails off at lag 9, so an AR or ARIMA model should be considered for this series.

The Partial Autocorrelation Function of the series would be a useful tool to determine order of the AR model, because the PACF for an AR(p) process cuts off after lag p. However because of the document posted online stating: "The textbook discusses partial autocorrelation functions as well. You do not have the statistical software for the partial autocorrelation function, and the discussion in the text is weak. You need not use the partial autocorrelation function for the student project." (building ARMIA models: a step by step guide, page 4, step 4, http://33771.hs2.instantasp.net/Topic7752.aspx), this step is skipped.

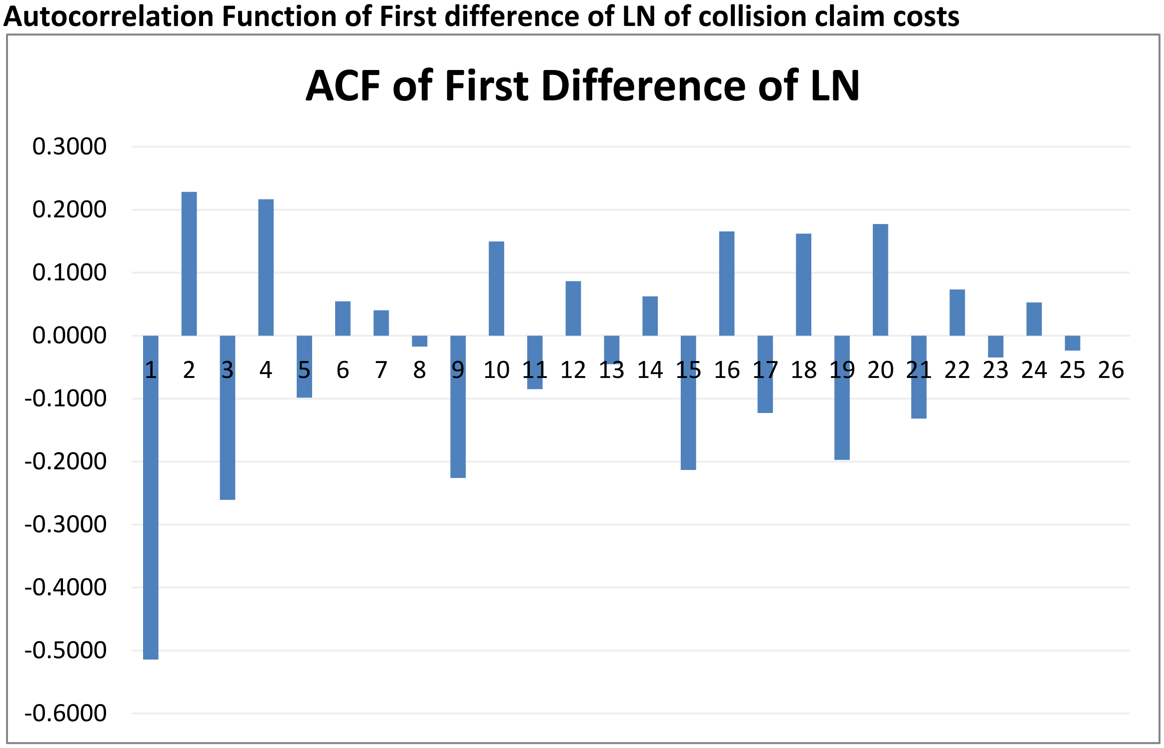
Instead, the first difference of the natural logarithm of collision claim cost is examined.



The first difference graph shows that values oscillate around zero, with no apparent pattern. The effect of trend has been removed, and the series is stationary now.

The Autocorrelation Function of the First difference of LN of collision claim cost is then looked at:

Autocorrelation Function of First difference of LN of collision claim costs



The ACF oscillates around zero, suggesting an AR model might be appropriate.

AR (1), AR (2) and AR(3) models are fitted and we will select the best fitted model.
Excel regression analysis tool was used to create the following results:

<u>AR(1) model</u>

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.87561349							
R Square	0.76669898							
Adjusted R Sc	0.75697811							
Standard Errc	0.09244898							
Observations	26							
ANOVA								
	df	*SS*	*MS*	*F*	*gnificance F*			
Regression	1	0.674	0.674099	79	5E-09			
Residual	24	0.205	0.008547					
Total	25	0.879						
	Coefficients	*ndard E*	*t Stat*	*-valu*	*ower 95%*	*Upper 95%*	*er 95*	*Upper 95.0%*
Intercept	1.26328511	0.795	1.588316	0	-0.378	2.90483	-0	2.904829772
X Variable 1	0.85019338	0.096	8.880957	0	0.6526	1.047775	1	1.047774893

The fitted AR(1) Model is Y_t= 1.26328511+0.85019338 Y_{t-1}

R square statistics is 0.76669898, meaning 76.7% of the variations of this series is explained by this AR(1) model.

Sum of coefficients for this model is 0.8501934, which is <1, hence this model is stationary.

AR(2) Model

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.906204618
R Square	0.821206809
Adjusted R Sc	0.804952883
Standard Errc	0.074955349
Observations	25

ANOVA

	df	SS	MS	F	ignificance F
Regression	2	0.57	0.283857	50.5236	5.97E-09
Residual	22	0.12	0.005618		
Total	24	0.69			

	Coefficients	dard E	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	Jpper 95.0%
Intercept	1.453095986	0.69	2.107295	0.046718	0.023047	2.883145	0.023047	2.883145
X Variable 1	0.371850928	0.17	2.224699	0.036673	0.02521	0.718492	0.02521	0.718492
X Variable 2	0.457330997	0.16	2.842531	0.009473	0.123668	0.790994	0.123668	0.790994

The fitted AR(2) Model is Y_t= 1.453095986+0.371850928 Y_{t-1} +0.457330997Y_{t-2}

R square statistics is 0.821206809, meaning 82.1% of the variations of this series is explained by this AR(2) model.

Sum of coefficients for this model is 0.829181925, which is <1, and each of the coefficients are less than 1 as well, hence this model is stationary.

AR(3) Model

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.889869
R Square	0.791867
Adjusted R Square	0.760647
Standard Error	0.077812
Observations	24

ANOVA

	df	*SS*	*MS*	*F*	*ignificance F*
Regression	3	0.460719	0.153573	25.36412	5.08E-07
Residual	20	0.121095	0.006055		
Total	23	0.581814			

	Coefficients	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Jpper 95.0%*
Intercept	1.670804	0.799341	2.090228	0.049576	0.003409	3.3382	0.00340908	3.3382
X Variable 1	0.293564	0.222248	1.320883	0.201455	-0.17004	0.757165	-0.1700376	0.757165
X Variable 2	0.45613	0.192961	2.363848	0.028319	0.053621	0.858639	0.05362082	0.858639
X Variable 3	0.053832	0.196053	0.27458	0.786455	-0.35513	0.462792	-0.3551276	0.462792

The fitted AR(3) Model is $Y_t = 1.670804+0.293564\, Y_{t-1} +0.45613 Y_{t-2}+0.053832 Y_{t-3}$

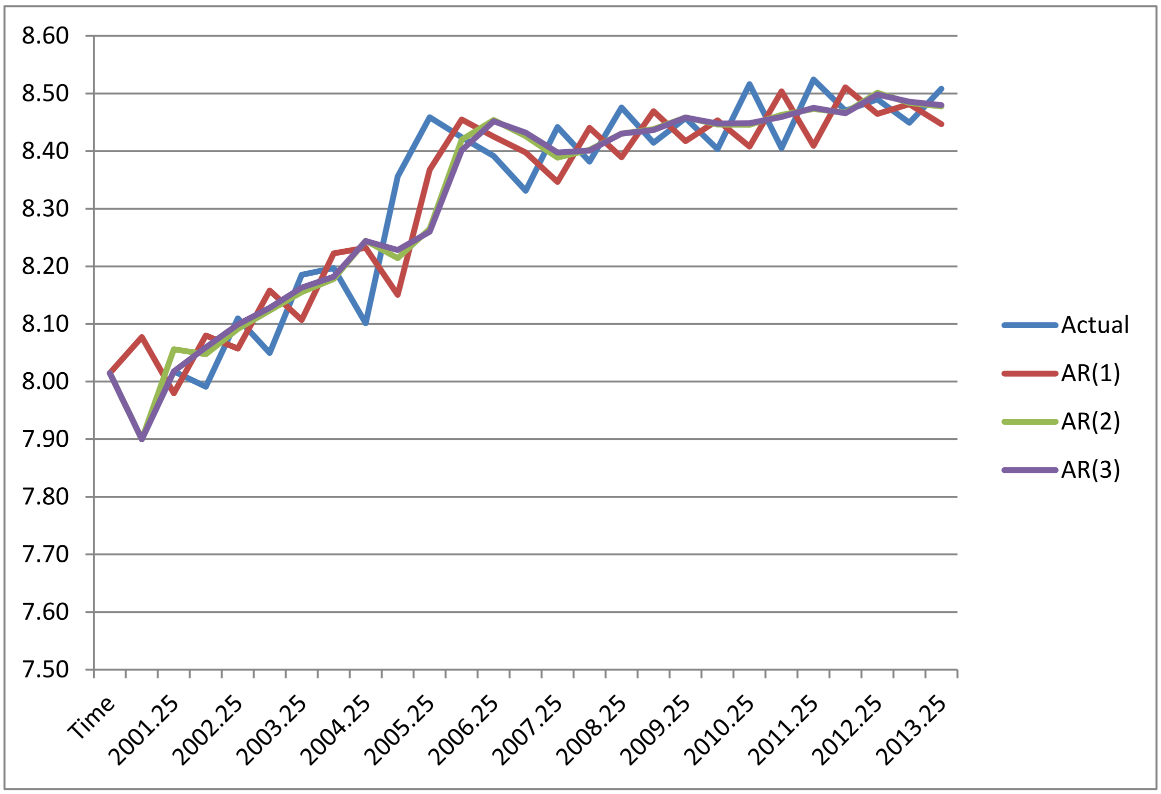
R square statistics is 0.791867, meaning 79.2% of the variations of this series is explained by this AR(3) model.

Sum of coefficients for this model is 0.509962, which is <1, and each of the coefficients are less than 1 as well, hence this model is stationary.

4) Actual vs. Fitted

Shown below is a graph of the natural logarithm of the actual data vs. the AR(1), AR(2) and AR(3) models:

Natural Logarithm of Actual vs. Fitted



5) Model Selection

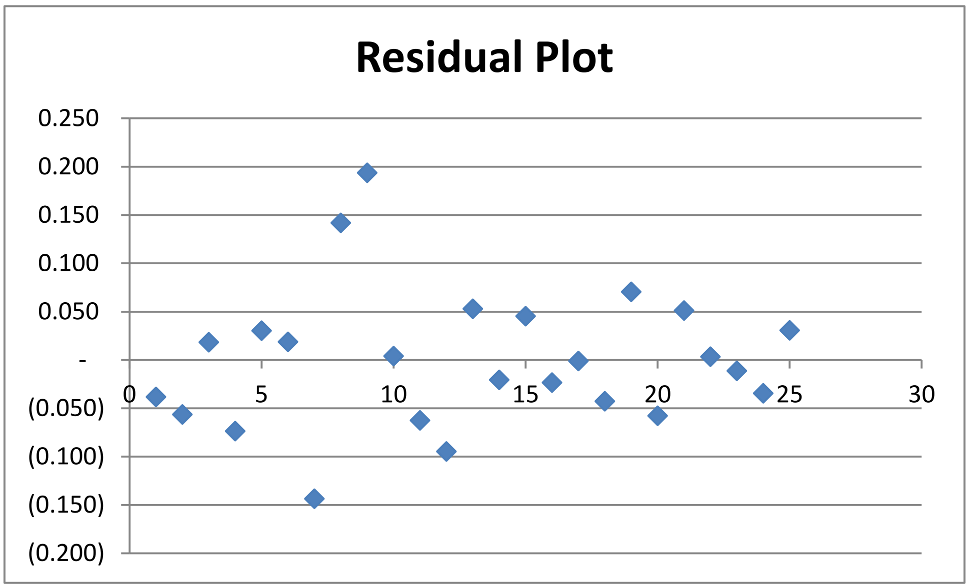
Based on actual vs. fitted figure above, and given the R square statistics for the three models ranges from 0.77 to 0.82, all models provide good estimates for the New Brunswick Automobile Industry collision claims cost data. But because of the higher R square statistics for AR(2), the AR(2) model is recommended.

The fitted AR(2) Model is Y_t= 1.453095986+0.371850928 Y_{t-1} +0.457330997$Y_{t-2.}$

6) Residuals

We will now look at the residual plot for the AR(2) model to further determine if the model is a good fit.

The residual plot for AR(2) model doesn't show any significant pattern, and the residuals oscilliates around zero, suggesting the model is appropriate.



The Durbin Watson statistics is performed to check for serial correlation. The Durbin Watson Statistic is greather than 2, serial correlation of the residuals is not present, further suggesting the model is appropriate.

RESIDUAL OUTPUT					
			Durbin Watson Statistic		
Observation	*Predicted Y*	*Residuals*	Residual Sq'ed	Residual Diff	Res Diff Sq'ed
1	8.056	(0.038)	0.001		
2	8.047	(0.056)	0.003	(0.018)	0.000
3	8.091	0.018	0.000	0.075	0.006
4	8.123	(0.074)	0.005	(0.092)	0.008
5	8.155	0.030	0.001	0.104	0.011
6	8.178	0.019	0.000	(0.012)	0.000
7	8.245	(0.144)	0.021	(0.162)	0.026
8	8.214	0.142	0.020	0.285	0.081
9	8.265	0.194	0.037	0.052	0.003
10	8.420	0.004	0.000	(0.190)	0.036
11	8.454	(0.063)	0.004	(0.067)	0.004
12	8.426	(0.095)	0.009	(0.032)	0.001
13	8.389	0.053	0.003	0.148	0.022
14	8.402	(0.021)	0.000	(0.074)	0.005
15	8.430	0.045	0.002	0.066	0.004
16	8.438	(0.023)	0.001	(0.069)	0.005
17	8.458	(0.001)	0.000	0.022	0.000
18	8.446	(0.043)	0.002	(0.042)	0.002
19	8.446	0.070	0.005	0.113	0.013
20	8.463	(0.058)	0.003	(0.128)	0.016
21	8.473	0.051	0.003	0.109	0.012
22	8.467	0.003	0.000	(0.048)	0.002
23	8.501	(0.011)	0.000	(0.015)	0.000
24	8.484	(0.034)	0.001	(0.023)	0.001
25	8.478	0.031	0.001	0.065	0.004
		Totals	0.124	0.069	0.264
			Durbin Watson Statistic		2.136

7) Conclusions

The AR(2) model is the suggested model for the New Brunswick Automobile Industry Collision Claim Cost time series after examining all of the above statistics.