Stochastic Trends in Policy Movement

A Time Series Analysis Project

Exploring the historic movements of inforce policies and new business of an Insurance Company to develop a predictive model.

Luanna C. Rahman
4/25/2014

INTRODUCTION

An insurance company has varying movements in its policies counts over the years. Whilst new issues represent a portion of the movement in policy count, its in force business at the beginning of each calendar year is subject to various decrements such as death, surrenders, lapses and expiries.

The most significant causes of these movements are due to new issues, death and surrenders. The insurance company models and projects death utilizing a US based mortality table. For surrenders, the company reserves a flat percentage of in force business. New business is modeled by sales targets for the forth coming year which from prior exercises have not provided a reasonable estimation.

Through analyzing historical data, this exercise is to determine if there exist any stochastic trends that can be modeled as a forecasting tool to reasonably project new business and surrenders of the insurance company to provide more accurate estimations than currently being used.

SOURCE OF DATA

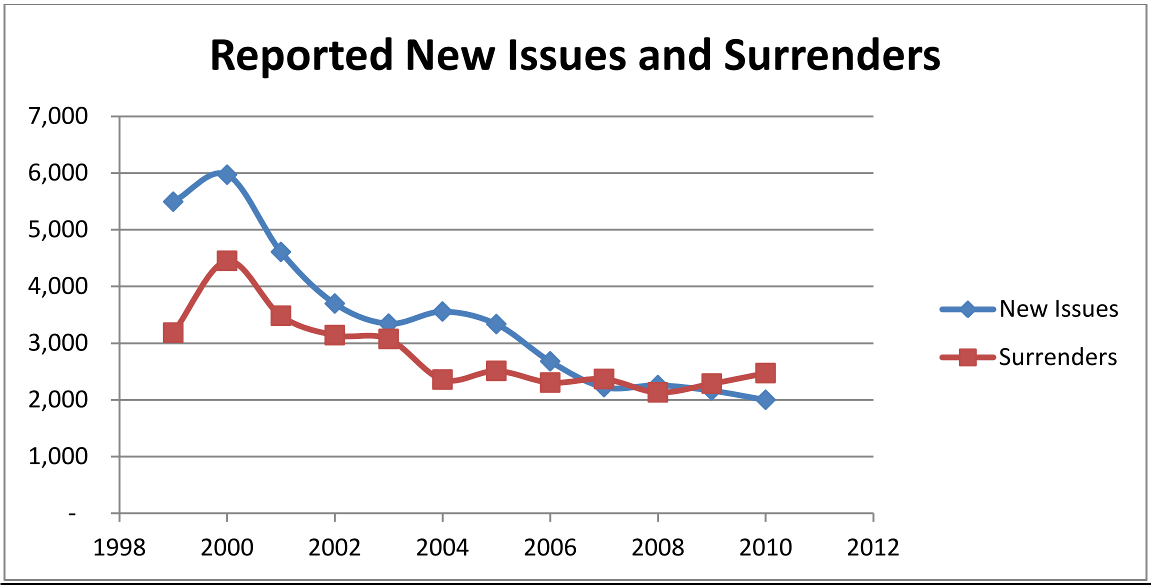
Monthly historic data from the company was gathered for the period January 1999 to December 2010. Prior to this 12 year period limited data was available and for some years monthly details could not be derived. Prior to this period, the company also closed some of its line of business and launched new products which would have affected these activities. The 12 year period selected represents a stable period for the company where the products offered were relatively consistent throughout the period.

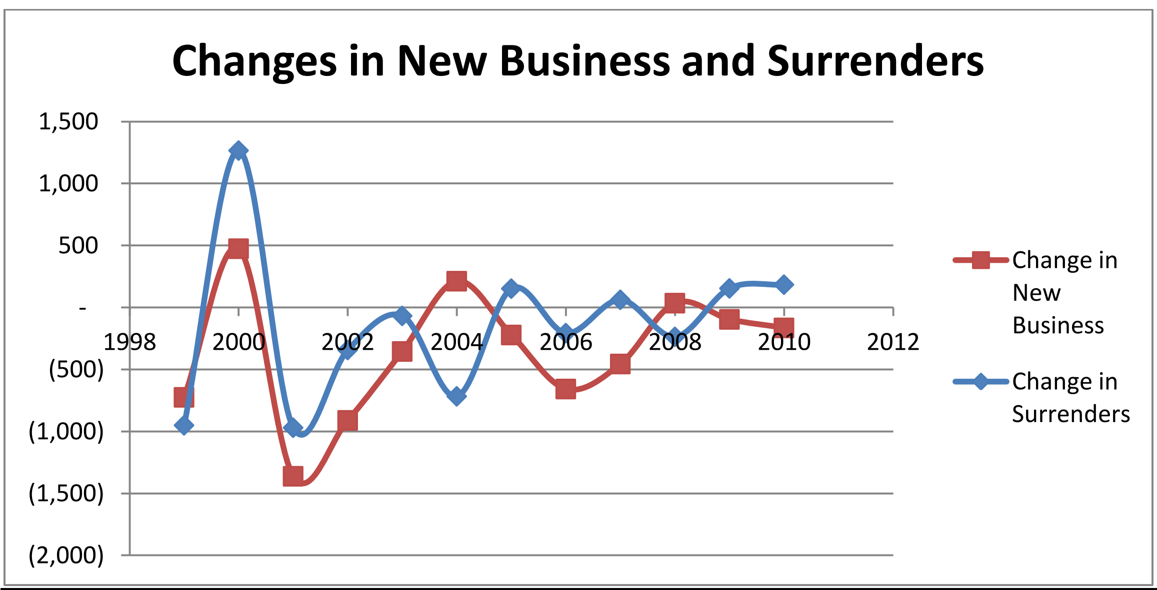
DATA OBSERVATIONS

The overall decline in New Issues over the twelve year period (as seen in the graph below) illustrates non-stationarity of the data series. Stationarity is evident in the changes in the new business activity. Also, the changes in new business illustrate relatively more volatility during the first half of the analysis period than evident in the latter.

The twelve year series of Surrenders also shows a similar though not as significant trend as the New Business series. Also in most cases consecutive data points tend to be closely related. The series of changes in surrenders also illustrates more volatility during the first four year.

The series of changes shows more stationary random movements.





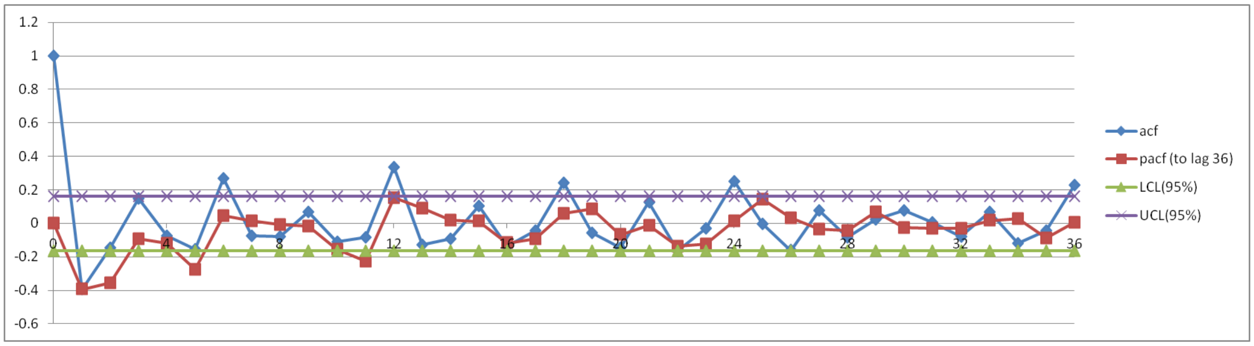
Monthly movements were then observed to determine if there were any apparent seasonality trends. From Appendix II – Graph C, there was no indication of such in either series. There did appear to be a correlation in consecutive points of the changes in New Business. There were also several observations of up to three consecutive points showing correlation dependencies in the series of changes in surrenders.

AUTOREGRESSIVE CORRELATIONS

Illustrations of the autocorrelations and partial autocorrelations were analyzed to provide further details of the autoregressive correlation characteristics of these series.

Changes in New Business

There are noted significant dependencies at time lags of 1, 6, 12, 18, and 24 in the graph of autocorrelation. The correlations of the latter time lags of 18 and 24 were just above the level of significance.

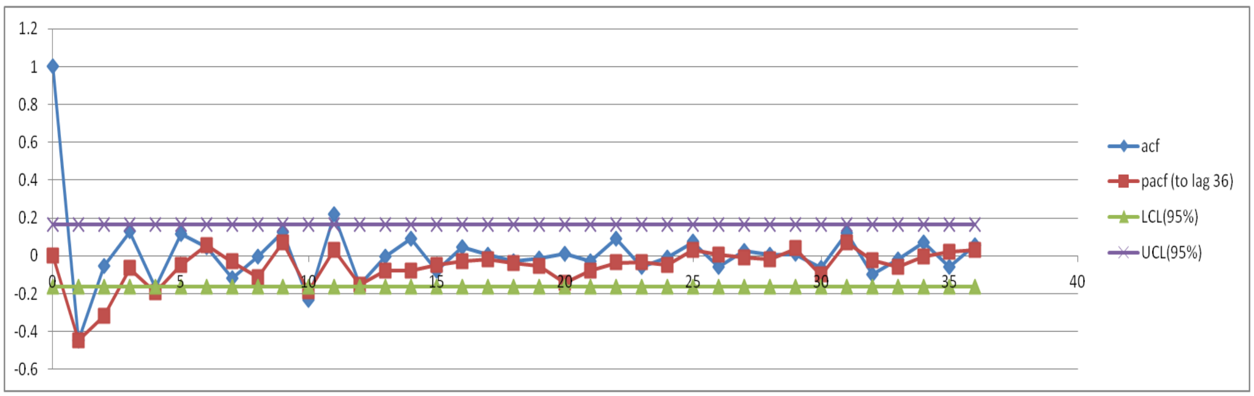


Partial autocorrelations tended to be close to zero for time lags greater than 12. There was a noted significant partial autocorrelation at time lag of 2. However as there are no reasonable grounds for consideration, the time lag of 2 would be ignored.

As there is evidence of seasonality in the correlations, time lags of 1, 6, and 12 would be considered in an Autoregressive model.

Changes in Surrenders

Only time lag of 1 has a significant autocorrelation while the partial autocorrelation function illustrates that time lags of 1 and 2 should be considered.



Therefore both AR (1) and AR (2) models would be explored as representations for this series.

DATA MODELLING

Data points were set as response variables against their time lagged counterpoints which were set a predictor variables to determine the required coefficients for the below models through regression.

A. Changes in New Business

Model I: $Y_t = \phi Y_{t-1} + \varepsilon_t$

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.421315966							
R Square	0.177507143							
Adjusted R Square	0.17167386							
Standard Error	82.06334206							
Observations	143							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	204927.9573	204927.9573	30.4300602	1.60573E-07			
Residual	141	949549.2874	6734.392109					
Total	142	1154477.245						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-2.454141131	6.865743854	-0.357447231	0.721291488	-16.02724614	11.11896388	-16.02724614	11.11896388
X Variable 1	-0.400340963	0.072573593	-5.516344823	1.60573E-07	-0.543813982	-0.256867945	-0.543813982	-0.256867945

Therefore $Y_t = -0.4003Y_{t-1} - 2.4514$

Model II: $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-6} + \varepsilon_t$

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.490258565							
R Square	0.24035346							
Adjusted R Square	0.229099438							
Standard Error	79.90015505							
Observations	138							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	2	272689.1311	136344.5656	21.3571151	8.73535E-09			
Residual	135	861844.695	6384.034777					
Total	137	1134533.826						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-2.440979149	6.807478392	-0.358573176	0.720474542	-15.9040764	11.0221181	-15.9040764	11.0221181
X Variable 1	-0.400547259	0.076575116	-5.230775722	6.28354E-07	-0.551989273	-0.249105245	-0.551989273	-0.249105245
X Variable 2	0.212885774	0.07253494	2.934941076	0.003922137	0.069433986	0.356337563	0.069433986	0.356337563

Therefore $Y_t = -0.4005Y_{t-1} + 0.2129Y_{t-6} -2.4410$

Model III: $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-6} + \phi_3 Y_{t-12} + \varepsilon_t$

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.559641777							
R Square	0.313198919							
Adjusted R Square	0.297102018							
Standard Error	76.7422722							
Observations	132							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	3	343770.4644	114590.1548	19.45709497	1.86346E-10			
Residual	128	753840.1719	5889.376343					
Total	131	1097610.636						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-2.890549606	6.687472075	-0.432233522	0.666298988	-16.12285548	10.34175627	-16.12285548	10.34175627
X Variable 1	-0.390944758	0.074359553	-5.257492044	5.94193E-07	-0.538077836	-0.24381168	-0.538077836	-0.24381168
X Variable 2	0.110605508	0.077936668	1.419171635	0.158279572	-0.043605502	0.264816518	-0.043605502	0.264816518
X Variable 3	0.284053017	0.0723345	3.926936901	0.000139851	0.140926853	0.427179182	0.140926853	0.427179182

Therefore $Y_t = -0.3909Y_{t-1} + 0.1106Y_{t-6} - + 0.2841Y_{t-12} - 2.8905$

Of note, all of the models meet the stationary conditions. All coefficients are between -1 and 1, and the sum of the coefficients is less than 1 for each model. However the models described, provide for only 17% to 30% of the variation of the series. As such, consideration is thus given to the previously excluded time lag of 2 to determine if its inclusion would increase the adjusted R Square goodness of fit measurement.

Model IV: $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-6} + \phi_4 Y_{t-12} + \varepsilon_t$

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.636861214							
R Square	0.405592205							
Adjusted R Square	0.3868707							
Standard Error	71.67447837							
Observations	132							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	4	445182.3185	111295.5796	21.66450816	1.20718E-13			
Residual	127	652428.3178	5137.230849					
Total	131	1097610.636						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-4.213821673	6.252951506	-0.673893228	0.501603882	-16.5872841	8.159640753	-16.5872841	8.159640753
X Variable 1	-0.544878841	0.077611429	-7.020600547	1.18549E-10	-0.69845785	-0.391299831	-0.69845785	-0.391299831
X Variable 2	-0.346107387	0.07789885	-4.443035881	1.90937E-05	-0.500255152	-0.191959623	-0.500255152	-0.191959623
X Variable 3	0.074062137	0.073253208	1.011042912	0.313918271	-0.070892741	0.219017014	-0.070892741	0.219017014
X Variable 4	0.242512014	0.068201689	3.555806588	0.000530218	0.107553182	0.377470846	0.107553182	0.377470846

Therefore $Y_t = -0.5449Y_{t-1} - 0.3461Y_{t-2} + 0.0741Y_{t-6} + 0.2425\, Y_{t-12} - 4.2138$

The model produced an increased Adjusted R Square of approximately 39% making the model with the best goodness of fit. The residuals generated by this model are provided in Appendix V.

B. Changes in Policies Surrendered

Model I: $Y_t = \phi\, Y_{t-1} + \varepsilon_t$

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.448904162							
R Square	0.201514947							
Adjusted R Square	0.195851932							
Standard Error	92.69555128							
Observations	143							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	305757.6758	305757.6758	35.5843949	1.8746E-08			
Residual	141	1211537.597	8592.465227					
Total	142	1517295.273						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-0.769173667	7.752216734	-0.099219835	0.92110465	-16.0947748	14.55642746	-16.0947748	14.55642746
X Variable 1	-0.449154195	0.075294918	-5.965265702	1.8746E-08	-0.598007088	-0.300301302	-0.598007088	-0.300301302

Therefore $Y_t = -0.4492Y_{t-1} - 0.7692$

Model II: $Y_t = \phi_1\, Y_{t-1} + \phi_2\, Y_{t-2} + \varepsilon_t$

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.538988542							
R Square	0.290508649							
Adjusted R Square	0.28030014							
Standard Error	87.87327048							
Observations	142							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	2	439481.2405	219740.6203	28.45750137	4.37345E-11			
Residual	139	1073317.921	7721.711665					
Total	141	1512799.162						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-0.588725118	7.375488948	-0.079821843	0.936493758	-15.17137735	13.99392712	-15.17137735	13.99392712
X Variable 1	-0.600708106	0.080056633	-7.50353943	6.72483E-12	-0.758994296	-0.442421916	-0.758994296	-0.442421916
X Variable 2	-0.32539532	0.079909685	-4.07203858	7.78832E-05	-0.483390966	-0.167399673	-0.483390966	-0.167399673

Therefore $Y_t = -0.6007Y_{t-1} - 0.3254Y_{t-2} - 0.5887$

Lag 10 was previously not considered as it lied on the upper level of significance. However its inclusion increases the goodness of fit measurement as seen below in Model III.

Model III: $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-10} + \varepsilon_t$

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.555015719							
R Square	0.308042448							
Adjusted R Square	0.292074197							
Standard Error	86.01504415							
Observations	134							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	3	428176.9639	142725.6546	19.29093201	2.08011E-10			
Residual	130	961816.4167	7398.587821					
Total	133	1389993.381						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-0.623837668	7.432961317	-0.08392855	0.933242344	-15.32906249	14.08138715	-15.32906249	14.08138715
X Variable 1	-0.545442127	0.081231773	-6.714640148	5.30426E-10	-0.706149473	-0.38473478	-0.706149473	-0.38473478
X Variable 2	-0.290209931	0.079791136	-3.637119942	0.000396197	-0.448067148	-0.132352713	-0.448067148	-0.132352713
X Variable 3	-0.189641616	0.074123685	-2.558448302	0.011660429	-0.33628646	-0.042996772	-0.33628646	-0.042996772

Therefore $Y_t = -0.5454Y_{t-1} - 0.2902Y_{t-2} - 0.1896\ Y_{t-10} - 0.6238$

This model, as did the prior two, meets the stationarity criteria, and as it has the highest Adjusted R Square measurement is considered the model of choice. The residuals generated by this model are provided in Appendix VI.

ANALYSIS OF MODEL RESIDUALS

New Business

There was more volatility in the model residuals at the beginning of the period of analysis. The range of errors was over 450 during this period. The range reduced by more than 50% during the last four years of analysis.

Surrenders

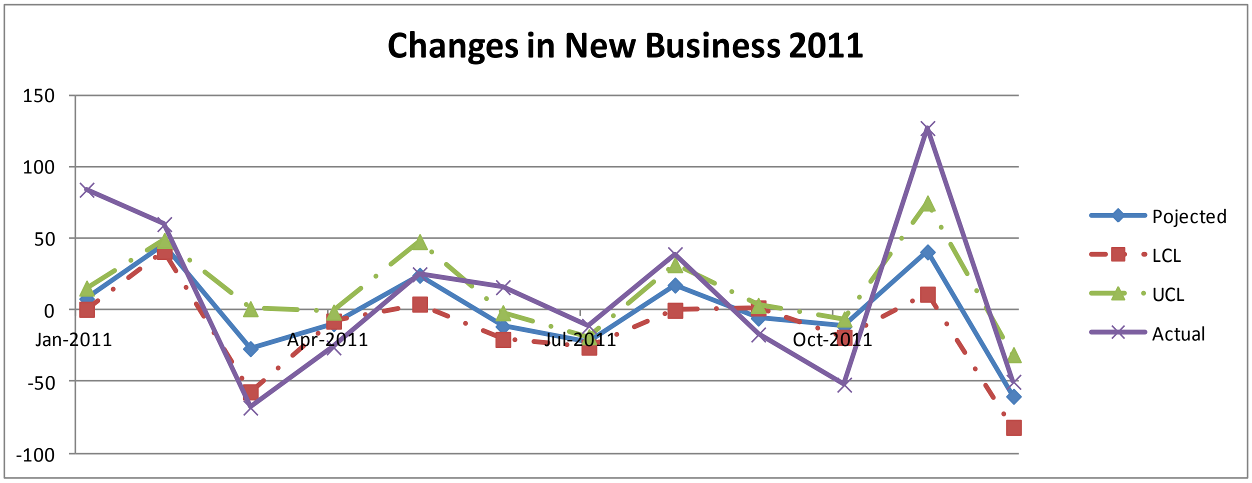
As was observed with the Changes in New Business series, there was noted significant volatility at the beginning of the series. There were also patterns in the residuals throughout the series.

MODEL PROJECTIONS

Using the models that provided the optimal goodness of fit measure, a twelve month forecast for each series was undertaken and compared to actual changes recorded during the period.

Changes in New Business

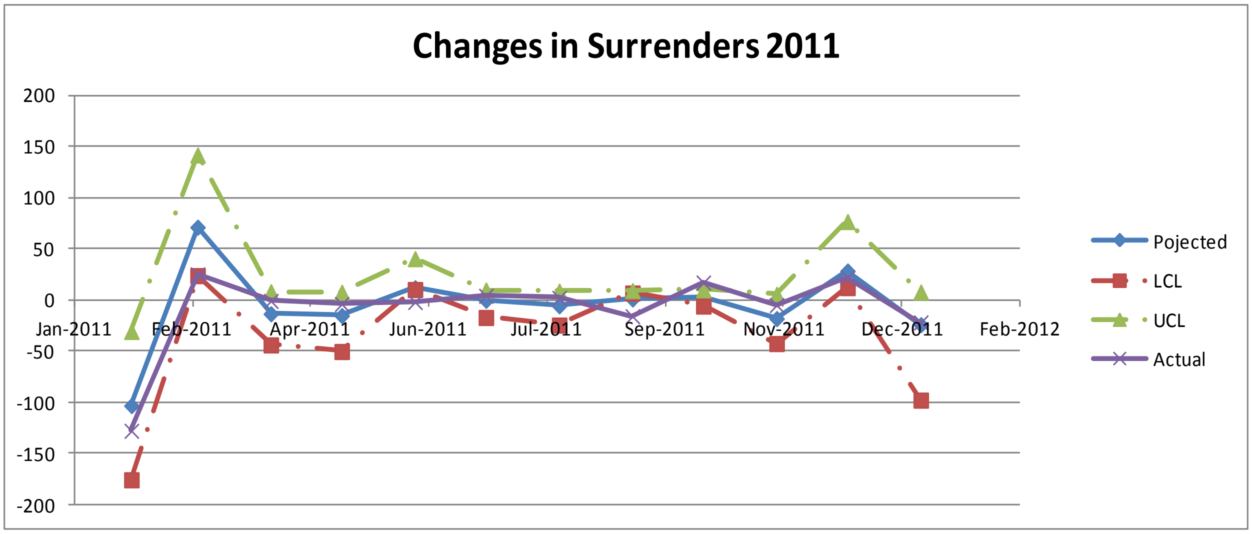
Model Selected: $Y_t = -0.5449Y_{t-1} - 0.3461Y_{t-2} + 0.0741Y_{t-6} + 0.2425\ Y_{t-12} – 4.2138$



There were some months where actual changes that were within the 95% Confidence Interval forecasted. However the actual changes for the majority of the year were outside the projected Confidence Interval. Three month had significant deviations from that expected.

<u>Changes in Surrenders</u>

Model Selected: $Y_t = -0.5454Y_{t-1} - 0.2902Y_{t-2} - 0.1896\ Y_{t-10} - 0.6238$



In general, the model produced accurate forecasts of the changes in surrenders. With only three exceptions, actual changes recorded during the year were within the projected 95% Confidence Interval.

CONCLUSION

Only the model for changes in policies surrendered provided to be a reasonable tool for a one year projection of the activity. As the models developed did not account for even half of variation in the two data series, it is anticipated that further forecasts would deviate further from the actual experience.

Further modeling should be undertaken to develop these models further so that they provide a better fit to the data series. It should be explored that ARIMA models may better represent the series than the Autoregressive models defined in this exercise. Industry, market or country economic indicators may account for the residuals produced by the Autoregressive models.

Public access to the Economic indicators for this region is limited and the required monthly statistical detail would have to be requested from government agencies, statistical institutions and research institutions.

APPENDICES

Data, tables and illustrations references in this exercise are provided in the attached file.