NEAS VEE – Time Series Project

Forecasting the Growth of Insurance Markets (Premiums) in Korea

Sunggon Jang

Summer 2015

A) INTRODUCTION

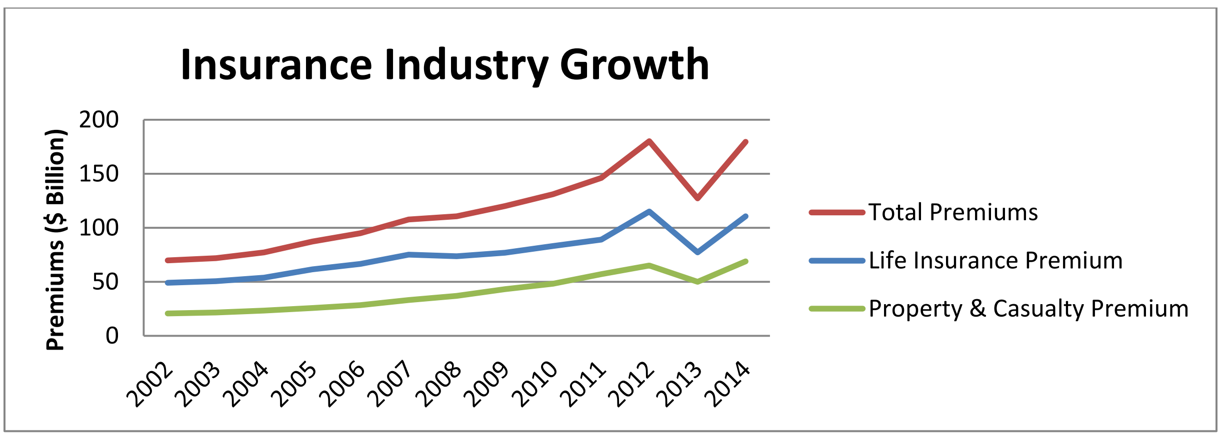
The idea of insurance started back in early human society, and the first methods of transferring or distributing risk in a monetary economy were practiced by Chinese and Babylonian traders in the 3rd and 2nd millennia B.C. respectively. Then, insurance became more sophisticated in Europe, and some forms of insurance were developed in London in the early decades of 17th century. In Korea, the first insurance company was established in 1946 right after the World War II, and since then the size of insurance industry has been growing rapidly. Since the insurance industry is closely related to the job markets for actuaries, we will do some research on the historic insurance premium in Korea and build a time series model for the overall growth of insurance premiums.

B) DATA

In this project, we will use the datasets from Korea Financial Supervisory Service, and we will look at the overall premium of Life insurance and P & C insurance industries from 2002 to 2014. The datasets can be downloaded from the following website: http://english.fss.or.kr/fss/en/main.jsp

C) ANALYSES

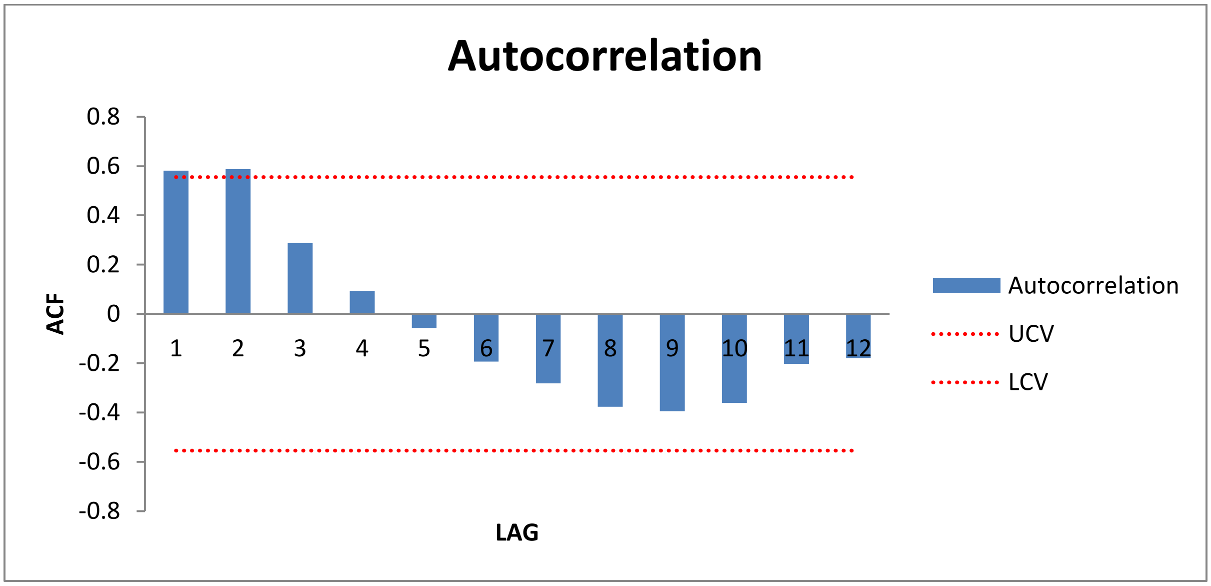
First of all, we will look at the overall growth of insurance markets in Korea from 2002 to 2014.



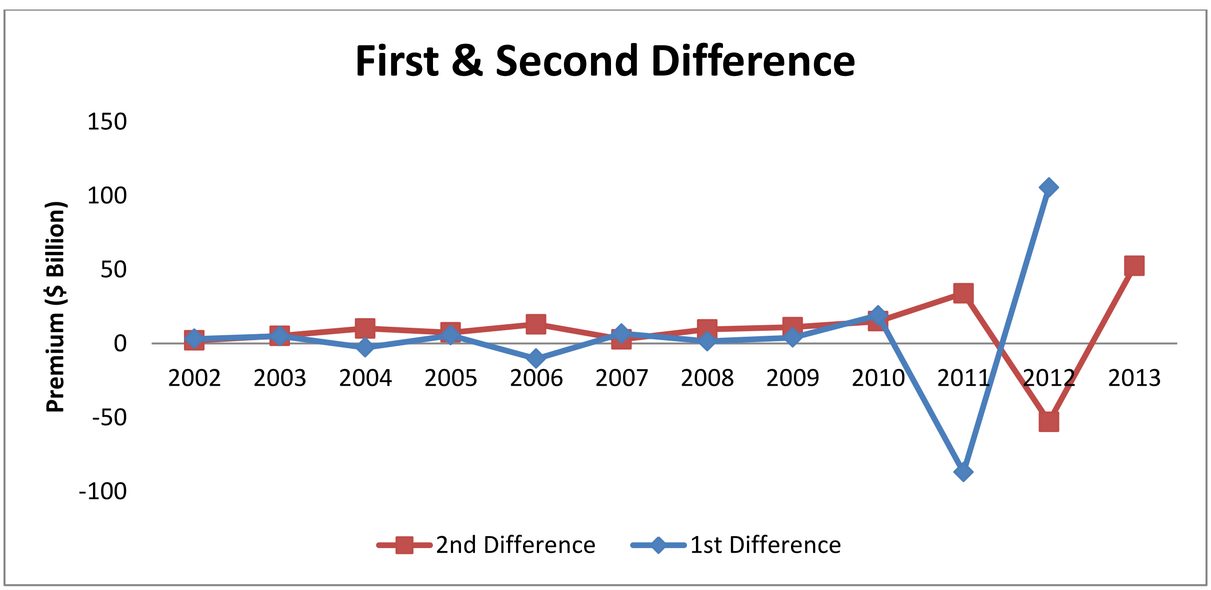
By looking at the past 12 years of data of insurance premium, we can observe that the overall insurance market has been growing in Korea (except that there was a downward trend from 2012 to 2013). Despite the trend pattern showing above, the process could not be a stationary. Therefore, in order to test this hypothesis, we can compute the sample autocorrelation function (ACF) at difference lags using the following formula:

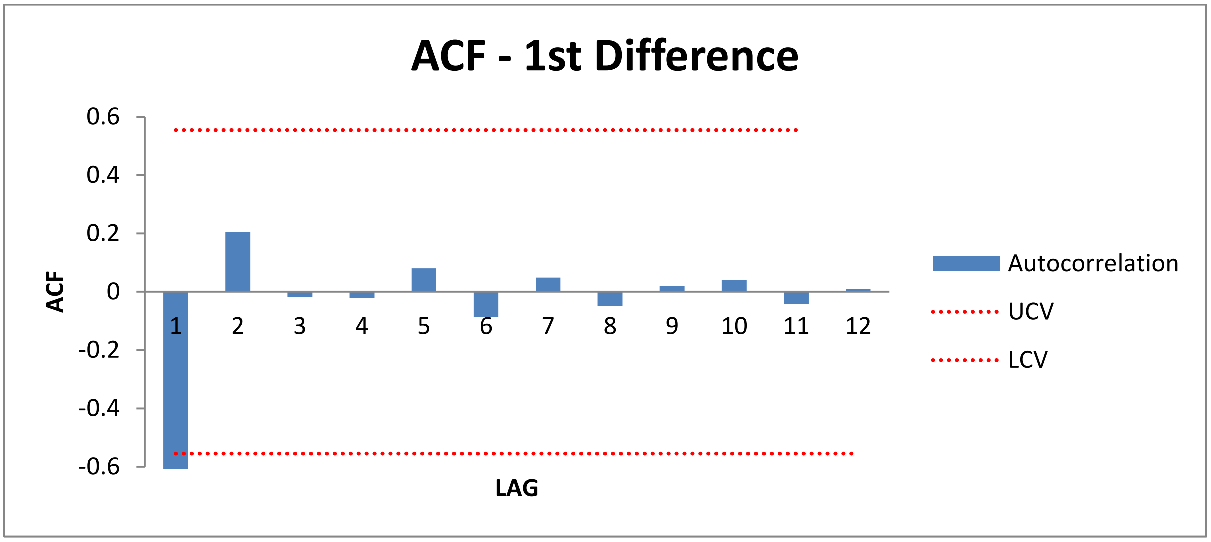
$$r_k = \frac{\sum_{t=k+1}^{n} (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2} \qquad \text{for } k = 1, 2, \ldots$$

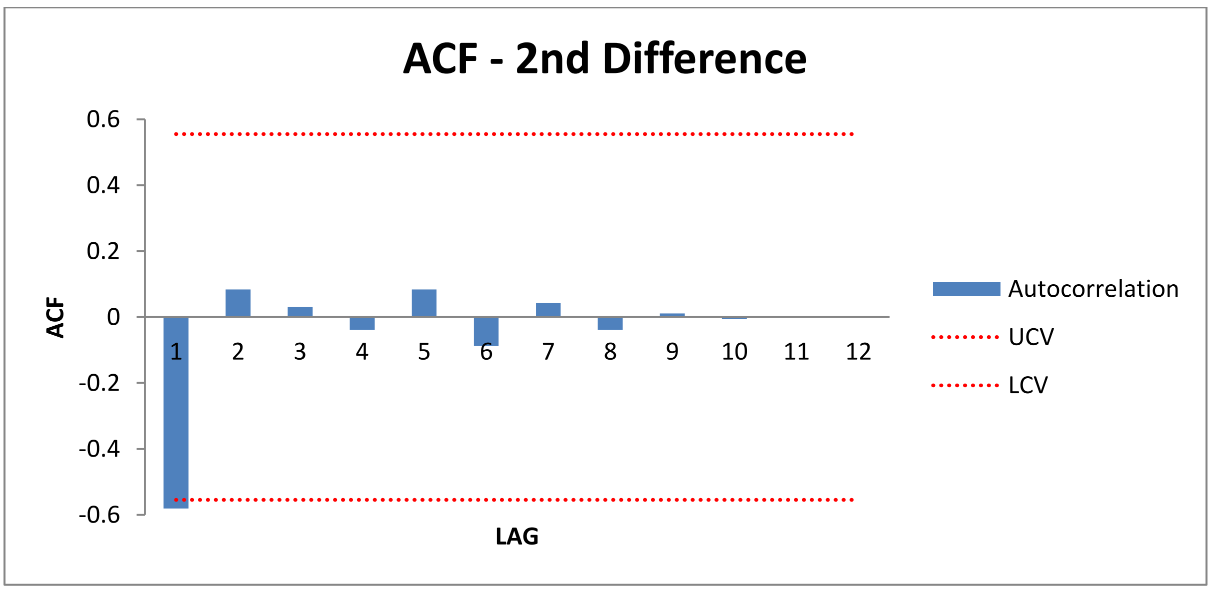
The autocorrelation graph below shows that the autocorrelation starts high at lag 1 and slowly decreases, reaching zero about at lag 4, and starts to increase again. Since the autocorrelations do not reduce to zero quickly enough, it is possible that this pattern represents an autoregressive, AR (p), model (ex. exponentially decays to 0 as the lag increases). Therefore, we can choose either the AR (1) or AR (2) process and test which model could be more suitable. Also, we can test out the first and second difference to verify that this process is a stationary model.



D) FIRST & SECOND DIFFERENCE & ACF







The graphs of fluctuation of the price of the first & second difference suggest that the process is stationarity. Moreover, the ACF oscillates around zero, suggesting an AR (P) model might be appropriate.

E) MODEL FITTING AND DIAGNOSIS – AR (1) vs AR (2)

Now, we will use excel regression analysis to fit the data to the following AR (1) model:

AR (1): $Y_t = e_t + \phi_1 * Y_{t-1}$

Summary						
Regression Statistics						
Multiple R	0.7524					
R Square	0.5661					
Adjusted R Square	0.5227					
Standard Error	24.7439					
Observations	12					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	1	7988.5549	7988.5549	13.0477	0.0048	
Residual	10	6122.5918	612.2592			
Total	11	14111.1467				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Y-Intercept	29.279	25.943	1.129	0.285	-28.526	87.083
Yt-1	0.817	0.226	3.612	0.005	0.313	1.322

From the analyses above, the fitted AR (1) Model is the following: **Yt = 29.279 + 0.817*Yt-1**

The R square for this model is approximately 0.5661, meaning 56.6% of the variations of this time series is explained by this AR (1) model, and$|\Phi_1|$= 0.817, which is <1; hence this proves again that the model is stationary.

Next, we will re-run the regression analysis and test the AR (2) model:

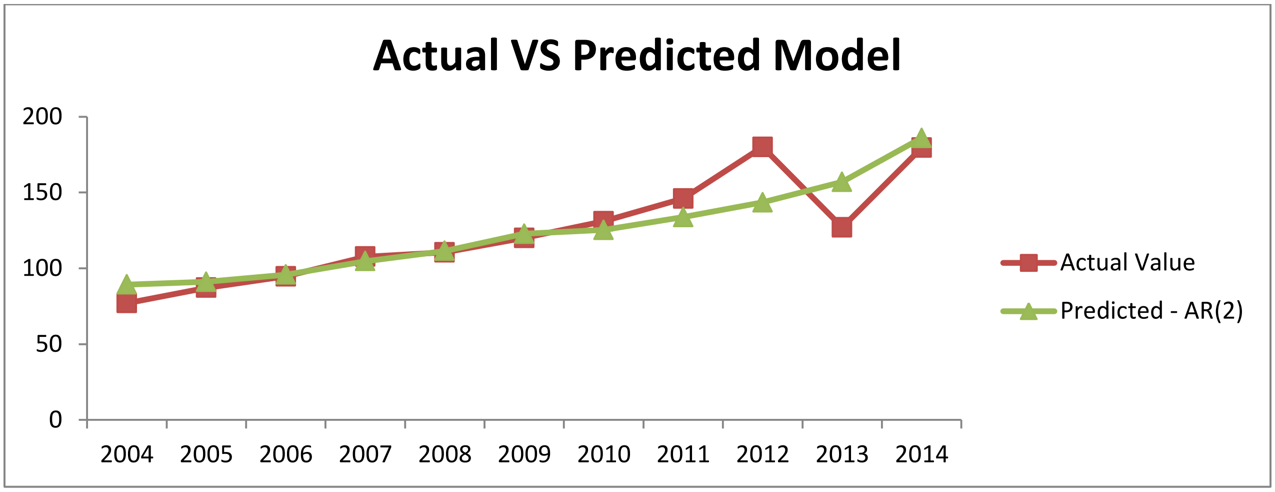
AR (2): $Y_t = e_t + \phi_1 * Y_{t-1} + \phi_2 * Y_{t-2}$

Regression Statistics						
Multiple R	0.87945829					
R Square	0.773446884					
Adjusted R Square	0.716808605					
Standard Error	18.15808454					
Observations	11					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	2	9005.140818	4502.570409	13.65590369	0.002634391	
Residual	8	2637.728273	329.7160342			
Total	10	11642.86909				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	27.72048982	21.47151475	1.291035595	0.232746159	-21.79291199	77.23389163
Yt-1	0.012533118	0.298099053	0.042043466	0.967494439	-0.674884532	0.699950768
Yt-2	0.86983261	0.278472414	3.123586274	0.014151765	0.227674072	1.511991149

From the analyses above, the fitted AR (2) Model is the following:

Yt = 27.7204 + 0.01253*Yt-1 + 0.8698*Yt-2

Comparing the first AR (1) model to AR (2) model, the adjusted R-squared value is much higher on the second model. However, because a large amount of premium declined in the year of 2012 to 2013, the adjusted R-square value is lower than expected; therefore, these datasets are somewhat hard to establish as a forecasting model.



F) Durbin-Watson Statistic

Finally, let's test the Durbin-Watson statistic to detect the presence of autocorrelation in the residuals from a regression analysis. The value of d always lies between 0 and 4, and if the value is close to 2, it indicates that there is no autocorrelation. Simply, the size of the residual for one case has no impact on the size of the residual for the next case.

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2},$$

After doing the calculation on Excel (see attached), the d value for AR (1) is approximately at 2.73 and AR (2) is approximately at 2.33.

G) CONCLUSION

Based on the above datasets, the growth of insurance premium in Korea from 2002 to 2014 could be best modeled by an autoregressive model, AR (P). Because of the characteristics shown in autocorrelation function, we did not test the moving average model, MA (P).

Moreover, by using this forecasting model, we can expect how much overall the premium will be collected in the year 2015; *Yt = 27.7204 + 0.01253*Yt-1 + 0.8698*Yt-2.* After the calculation, it will be approximately about $140.4389 billion dollars of premium in year 2015.

However, the datasets we used are too small, and the period was too short to actually predict an accurate future values. For instance, there was a large premium decline in one year, and it is hard to be explained in the model. Therefore, we will have to update these results in order to have a much more accurate forecasting model. This could be established by dividing the datasets quarterly and checking the seasonality or any other trends in the data. However, it is glad to see the insurance industry is growing!