Time Series Student Project

I. Introduction

As a new employee of the group insurance division in our company, I am interested to know the estimates of premium income coming from new group business sales. This project aims to find the model that would help the company project future sales from new group businesses using past data.

Total sales come from both group life business and group health business. It is assumed that there are no significant changes in the allocation of sales to each line of business and other external factors that would affect the premium income of the company.

II. Data

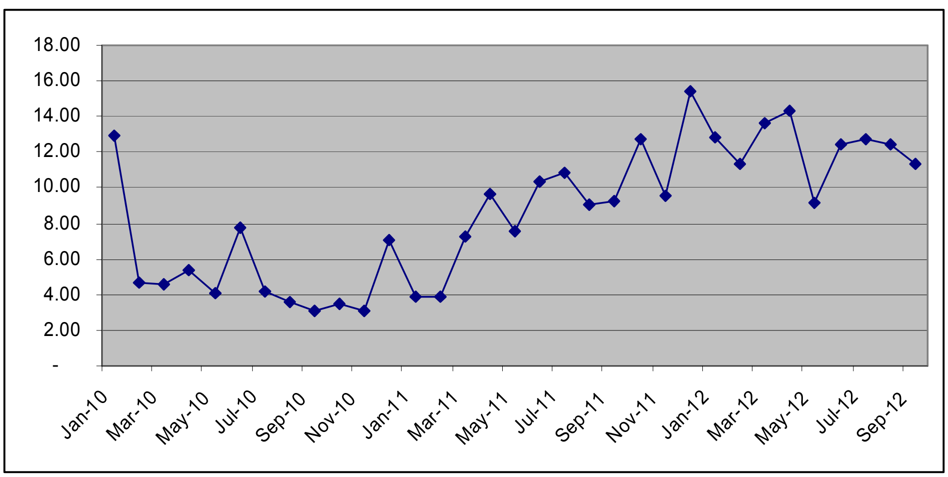
The data was collected monthly from January 2010 to September 2012 and was presented in millions (Philippine Pesos). The data used in this project can be seen in the attached Excel file in worksheet "Data".

Time Series
Project.xlsx

III. Analysis

First, the graph of the monthly new business sales is shown in Figure 1 below.

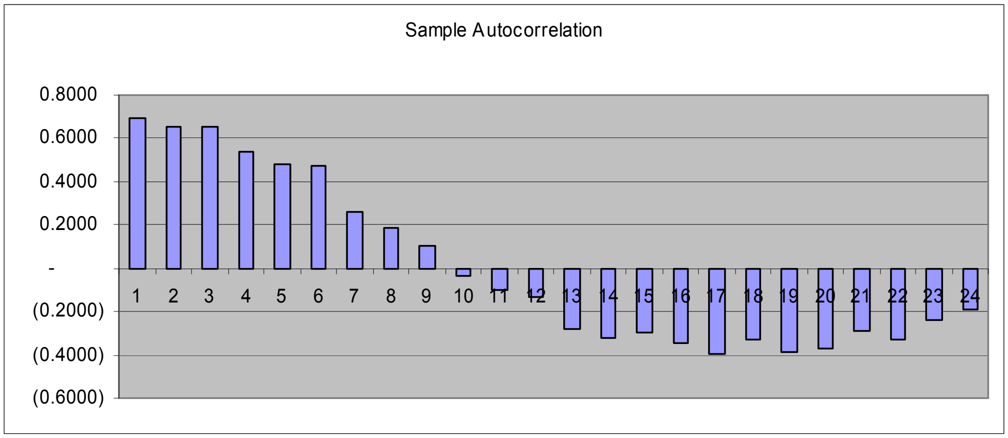
Figure 1



We can easily observe that there's no strong indication of seasonality in our data. To confirm this, the following chart in Figure 2 displays the sample autocorrelation against its lag. I used the formula below to graph the sample autocorrelation function. Please see attached excel file (in worksheet "Autocorrelation") to see the computations.

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

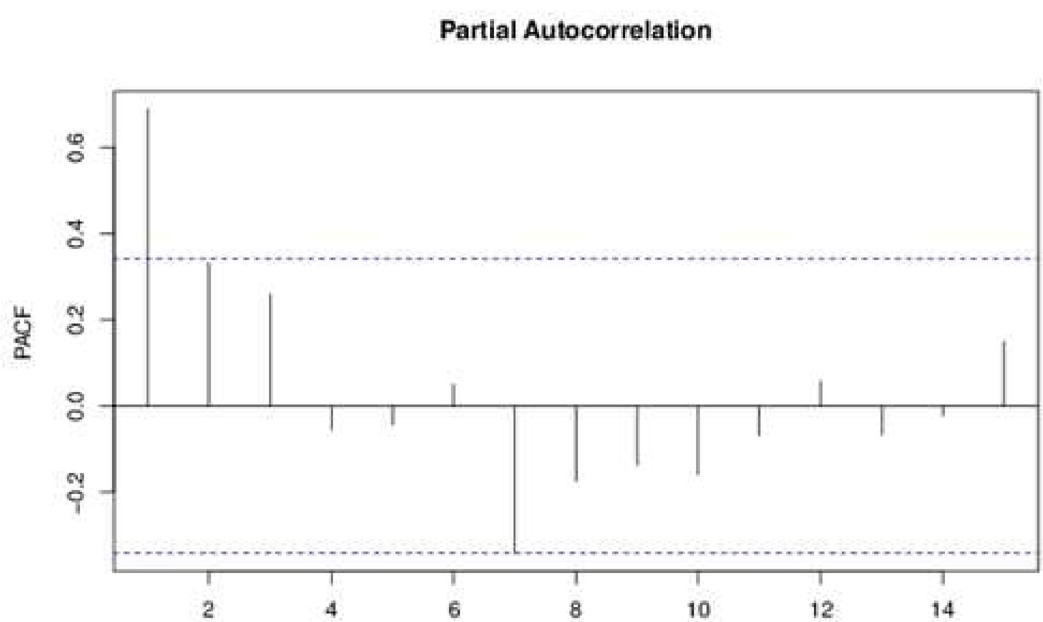
Figure 2



It can be seen that there are no evident peaks in our graph which would indicate seasonality. Furthermore, the ACF tails off at around lag 10. Based on this graph, we would certainly consider an AR or ARMA model for this time series.

We can compute for the sample partial autocorrelation to determine the order of autoregressive model and aid us in choosing the correct model. Please see Figure 3 for the partial autocorrelation plot against lag k.

Figure 3



We have learned that the partial autocorrelation function for an AR(p) process cuts off after lag p. Since our plot abruptly goes to almost 0 after lag 3 instead of tailing off, an AR(p) model would best fit this particular time series with p = 3. We can also consider an AR(1) or an AR(2) model for our data.

To help us choose the model that would best fit our time series, we use excel regression tool to get the coefficient parameters of the three AR models with p = 1, 2 and 3. The results are then given below:

A. For an AR(1) model, excel regression tool gives us the following:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.70954863
R Square	0.50345926
Adjusted R Square	0.4869079
Standard Error	2.74594668
Observations	32

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	229.3585297	229.4	30.418	5.44E-06
Residual	30	226.2066957	7.54		
Total	31	455.5652254			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	2.49399231	1.183112756	2.108	0.04349	0.0777537	4.91023089	0.077753723	4.910230893
X Variable 1	0.70051621	0.127014366	5.515	5.4E-06	0.4411183	0.95991415	0.441118274	0.959914155

Based on this, AR(1) model is $Y_t = 2.493992 + 0.700516Y_{t-1.}$

B. For an AR(2) model, excel regression tool gives us the following:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.80425727
R Square	0.64682976
Adjusted R Square	0.62160332
Standard Error	2.3574194
Observations	31

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	284.99526	142.5	25.6409	4.697E-07
Residual	28	155.6079343	5.557		
Total	30	440.6031943			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	1.41476107	1.093121113	1.294	0.20616	-0.824396	3.65391813	-0.824396	3.653918133
X Variable 1	0.57048624	0.156752319	3.639	0.00109	0.2493937	0.8915788	0.249393675	0.891578804
X Variable 2	0.28546708	0.155927954	1.831	0.0778	-0.0339368	0.60487101	-0.03393685	0.604871011

Based on this, AR(2) model is $Y_t = 1.414761 + 0.570486Y_{t-1} + 0.285467Y_{t-2.}$

C. For an AR(3) model, excel regression tool gives us the following:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.83827146
R Square	0.70269905
Adjusted R Square	0.66839509
Standard Error	2.20160441
Observations	30

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	297.8687803	99.29	20.4845	5.047E-07
Residual	26	126.0236115	4.847		
Total	29	423.8923918			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	1.04560546	1.057998373	0.988	0.33212	-1.1291413	3.22035224	-1.12914132	3.220352243
X Variable 1	0.32843801	0.176862328	1.857	0.07467	-0.0351077	0.69198373	-0.0351077	0.691983731
X Variable 2	0.31983966	0.178461131	1.792	0.08475	-0.0469924	0.68667176	-0.04699245	0.686671763
X Variable 3	0.27305882	0.155190293	1.76	0.09025	-0.0459394	0.59205703	-0.04593939	0.592057031

Based on this, AR(3) model is $Y_t = 1.045605 + 0.328438Y_{t-1} + 0.319840Y_{t-2} + 0.273058Y_{t-3.}$

To check if the three autoregressive models tested were stationary, we take the sum of the coefficients and see if they are each less than 1. We can see in Table 1 below that all models satisfy this condition and thus, the three models are stationary.

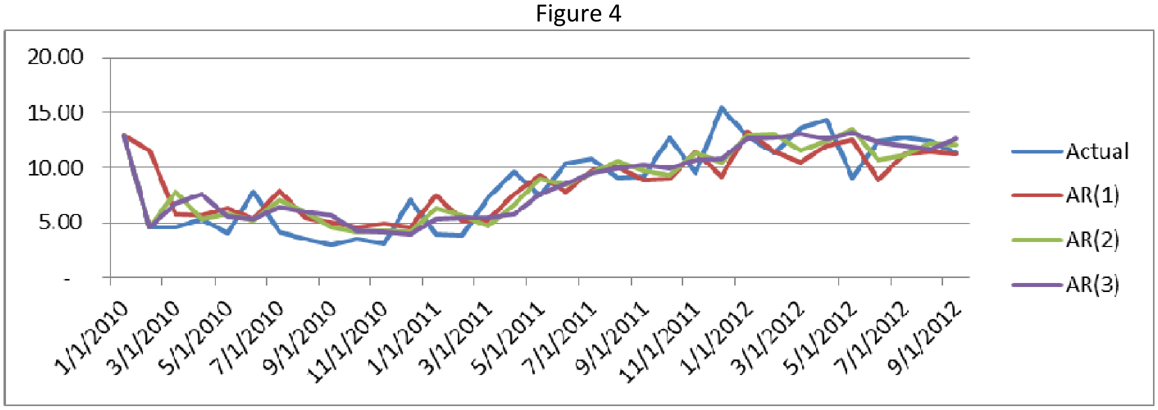
Table 1

Model	Sum of Coefficients	Less than 1?
AR(1)	0.700516215	<1
AR(2)	0.855953321	<1
AR(3)	0.921336491	<1

Also, the values of the coefficient of determination or multiple R-squared of the 3 models range from 50% to 70%. This means that around 50% to 70% of the variation in the time series is explained by the AR models.

We now plot the actual data against the AR models and decide which has the best fit.

Figure 4



IV. Conclusion

Based on Figure 4 above, all models provided good estimates for the new business sales. However, analysis of new business sales between January 2010 to September 2012 indicates that an AR(3) model is the best fit and is the recommended model to forecast new business sales for group insurance. We can now use our AR(3) model

$$Y_t = 1.045605 + 0.328438Y_{t-1} + 0.319840Y_{t-2} + 0.273058Y_{t-3}$$

to propose an updated monthly estimates in the remaining period of 2012 and the whole year of 2013.