

*Full Length Research Paper*

# Forecasting general insurance loss reserves in Egypt

Tarek Abd Elhamid Ahmed Taha<sup>1</sup>, Yusnidah Ibrahim<sup>2\*</sup> and Mohd Sobri Minai<sup>2</sup>

<sup>1</sup>College of Commerce, Tanta University, Egypt.

<sup>2</sup>College of Business, Universiti Utara Malaysia, Malaysia.

Accepted 1 June, 2011

**Loss reserve is one of the most important indicators that have many important and strategic decisions applications, such as rate making decisions, underwriting decisions, investment decision and corporate planning. The aim of this study is to identify the reliable time series forecasting model to forecast loss reserve estimates of Egyptian general insurance companies. Exponential smoothing model, Box-Jenkins analysis and time series regression model are applied on actual reported loss reserves data for general insurance sector for the period 1986 to 2006 and their accuracy are compared based on several error measures. The series from 1986 to 2001 are used for the estimations process and the remaining observations are used to evaluate the models as outside sample data. Exponential smoothing technique in all steps-ahead is identified as the best forecasting technique to Egyptian general insurance sector.**

**Key words:** General insurance, technical reserve estimation, time series forecasting.

## INTRODUCTION

One of the major tasks that insurance companies routinely perform is loss reserves estimation. Loss reserves, also referred to as claim reserves, are estimates of what claims will cost. The reserve represents money that is set aside to meet claims arising in the future on the policies already written (England and Verrall, 2002). In other words, the claim reserving process is a prospective estimation of what claims will cost to the insurance company.

The principal aim of reserving estimation is to insure the protection of policyholders (Cheung, 1997). Apart from that, loss reserving has important implications for insurers pricing and competitive responses. If the estimate of loss reserves were too low, premiums would be inadequate to support the financial projections of future periods. In the worst case scenario, rate would be insufficient to pay claims and the company would be insolvent, if the loss estimates were too high, consequently, insurance rates may be raised above competitive levels (Douglas, 1999; Calandro and O'Brien, 2004). Furthermore, general insurers need to be able to estimate loss reserves to make sure that they have sufficient assets to cover their liabilities. Insolvent insurance companies are

not allowed to continue to sell insurance policies because it does not have the financial strength to keep its contractual obligations to its policyholders (Cheung, 1997).

Loss reserves estimation are also important for outsiders in determining the financial liability, financial strength and value of an insurance company (Cheung, 1997). Loss reserves are collectively the largest liability on a property and liability insurer's balance sheet, thus any under-reserving of losses will decrease an insurer's liabilities and boost its net income (Grace and Lavery, 2006).

Loss reserves in general insurance market are estimated based on actuarial reserving techniques such chain ladder technique (CLT) (Harnek, 1966), Bornhuetter Ferguson technique (BFT) (Bornhuetter and Ferguson, 1972) and Taylor separation technique (TST) (Taylor, 1977). Nevertheless, data required by all these reserve estimation techniques is complex and ambiguous, where it is assumed that the claims paid can be grouped according to accident years and development years, resulting a claim run-off triangle (Panning, 2006).

Considering the importance of loss reserve figures for future planning and decision-making and the complexity and ambiguity of actuarial loss reserve calculation, it is beneficial to explore non-actuarial approaches to estimating loss reserves; one of them is the time series forecasting technique. Hence, this study aims is to

\*Corresponding author. E-mail: [yibrahim@uum.edu.my](mailto:yibrahim@uum.edu.my). Fax: 006049286406.

compare the accuracy of several time series forecasting techniques, namely exponential smoothing model, Box-Jenkins analysis and time series regression model, in order to identify the best forecasting model. To the authors' knowledge, there are only a few studies that explore the utilization of time series forecasting model in reserves estimation.

## PRIOR RELATED STUDIES

Studies that utilize time series technique to forecast loss reserve is scarce. Harbey (1995) used multiple regression technique to estimate the outstanding loss reserve in Egyptian general insurance market within the period from 1980 to 1994. He found that outstanding claim reserve estimation is affected by two variables, namely, underwriting premiums and paid claims, Harbey (1995) reached the possibility of applying the multiple regression models to some general insurance segments and could not be applied to others.

Doray (1996) considered the problem of forecasting the number of claims incurred. After subtracting the number of claims reported to date, the number of claims incurred but not reported (IBNR), which is the main portion of loss reserves, can be forecast. The basic model assumes that the number of claims per accident period follows an autoregressive moving average time series process. Instead of assuming the data available in the usual claim run-off triangle format, he assumed that the only data available are the number of claims reported at the valuation date for each accident interval of an observation period. Box-Jenkins methods are used to forecast the ultimate number of claims incurred and to obtain approximate confidence intervals for the number of claims incurred.

Chan et al. (2004) used the growing triangle technique (GT) for comparing between different models for loss reserve estimation including CLT which based on weighted average mean square error, CLT based on simple average mean, Bornhuetter- Ferguson, and log-normal model. At the base of GT technique, sub-triangle of losses, embedded in the full triangle of available data, are used to evaluate the prediction power of various candidate methods of estimation. The GT technique is illustrated using three data sets: i) paid losses during 1978 to 1995 ii) payments per claim during 1981 to 1995 iii) number of claims notified during 1985 to 1995. Based on weighted mean square error (WMSE) for the three run-off triangle, Chan et al. (2004) found that CLT based on the simple average provides the best prediction for the second and the third data sets. Also, Chan et al. (2004) found that the CLT technique produces relative constant results in the growing triangle assessment, this is explained by the relative simple and non-parametric nature. On the other hand, he found that the log-normal model is the best model for the first data set.

Bai et al. (2005) estimated loss reserves in the US

dental insurance based on traditional model which is the CLT and suggested two models namely log-linear regression model and time series model based on the holt-winter seasonal algorithm. Using data from General Motors employees provided by Delta Dental and using the CLT method on General Motors data from January 2001 to December 2002, they found that the traditional CLT model is viewed as being inadequate in the light of the error rates which are calculated based on the difference between total estimated reserve and the actual total reserve. Using the log-linear regression model on data from various corporate plans, and comparing the estimated values with the actual loss reserves, the results show that the variation in the error rates are very high and it is difficult to tell in advance if the prediction is going to be accurate. They also found that the model based on time series analysis using the Holt-Winters seasonal algorithm which is designed to handle both trend and seasonal variation in the data is performed well on many data sets by decreasing the error rates. They found that many techniques did not estimate dental insurance reserve with constant accuracy, some of suggested techniques performed well for certain data sets and performed poorly on others. They noticed that some of these models such as CLT and time series required 24 to 36 months of historical data while log-linear regression model required only 13 months data. So, for new insurer's, modeling with time series may not be possible. They concluded that there seems to be no model that will be both accurate and constant for all data sets.

## RESEARCH METHODOLOGY

### Forecasting models

Three time series methods are used to estimate loss reserves. These model forecasts are compared to the actual values of the data, and the forecasting models are ranked and compared depending on their standard error.

### Exponential smoothing technique

This technique is a most widely used class of univariate time series modeling technique and is a very popular scheme to produce a smoothed time series (Lazim, 2005). Whereas in moving averages, the past observations are weighted equally, exponential smoothing assigns exponentially decreasing weights as the observation gets older. In other words, recent observations are given relatively more weight in forecasting than the older observations.

In exponential smoothing models, the forecast for the next and all subsequent periods are determined by adjusting the current period forecast by apportion of the difference between the current forecast and the current actual value. In Brown's method, only one smoothing constant is used and as such, the estimated linear trend values obtained are sensitive to random influences. To overcome this problem, the study uses the Holt's method, which is a technique frequently used to handle data with linear trend. This method does not only smoothen the trend and the slope directly by using different smoothing constants, but also provides more flexibility in selecting the rates at which the trend and slopes are tracked. The application of Holt's method requires three equations: the

exponentially smoothed series:

$$s_t = \alpha y_t + (1 - \alpha)(s_{t-1} + T_{t-1})$$

$T_{t-1}$  The trend estimate, the trend estimate is calculated by taking the difference between two successive exponential smoothed values ( $S_t - S_{t-1}$ ):

$$T_t = \beta(s_t - s_{t-1}) + (1 - \beta)t - 1$$

A second smoothing constant  $\beta$  is used to smooth the trend estimate. In this equation, the estimate for the trend ( $S_t - S_{t-1}$ ) is multiplied by the smoothing constant  $\beta$ . This value is then incorporated into the previous estimate of the trend that has been adjusted by the factor  $(1 - \beta)$ . The smoothing is done for the trend rather than for the actual data, this result in a smoothed trend without any randomness:

$$F_{t+m} = s_t + T_t \times m$$

The values of  $\alpha$  and  $\beta$  are the parameters to be determined with values ranging from 0 to 1.

**BOX-Jenkins model (ARIMA)**

The ARIMA stands for the combination that comprises of autoregressive/integrated/moving average models. It is commonly applied to time series analysis, forecasting and control.

The basis of the Box-Jenkins / (ARIMA) modeling approach consists of three main stages: model identification; model estimation and validation; model application.

**Model identification:** The first step in the application of the Box-Jenkins methodology is to identify the class of model most suitable to be applied to the given data set. This is done by computing, analyzing and plotting various statistics based on historical data.

Common statistics used to identify the model type is the auto-correlation function (ACF), and the partial auto-correlation function (PACF). Hence, the common practice now is to identify several highly likely model formulations and subsequently choose the best model that meets all statistical requirements. The process of identifying the models is, thus, summarized as follows:

1. To compute and analyze the various statistics based on the historical data, in particular the ACF and the PACF.
2. Based on information obtained from 1 above, the most suitable subclass of the general class of model is then identified.

**Model estimation and validation:** The specific parameter values are estimated subject to the condition that the selected error measure is minimized. More specifically, the process is to search for the estimated parameter values that minimize the differences between the actual and the forecast values.

**Model application:** If all test criteria are met and the model's fitness has been confirmed, it is then ready to be used to generate the forecasts values. At this stage, three possibilities may occur; new or latest data are collected and incorporated into the existing

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Since the AR and the MA models are of order 'p' and 'q'

series; new model is formulated and re-estimated; develop a system to monitor the forecast values produced.

**Assumption of Box-Jenkins methodology:** The application of the Box-Jenkins methodology lies on the assumption that concerns the characteristic of the initial data series (Lazim, 2005). Basically, it is assumed that the data series is stationary. Where such assumption is not met, then the necessary procedures are performed in order to achieve stationary in the series. A simple procedure used to remove the non-stationary in time series is to perform the differencing, and log transformation is commonly used to stabilize the variance. There are four basic models:

- i. The autoregressive (AR) model,
- ii. The moving average (MA) model, and
- iii. Mixed autoregressive and moving average model.
- iv. Mixed autoregressive, Integrated and moving average model.

**i. The autoregressive (AR) model**

In the AR model, the current value of the variable is defined as a function of its previous values plus an error term (Lazim, 2005). Mathematically, it is written as:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

Where  $\mu$  and  $\phi_j$  (  $j = 1, 2, \dots, p$  ) are constant terms or parameters to be estimated,  $y_t$  : is the dependent or current value and  $y_{t-p}$  the  $p^{th}$ . order of the lagged dependent or current value, and  $\varepsilon_t$  : is the error with mean=0 and variance  $\sigma_e^2$ .

**ii. The moving average model (MA) model**

The moving average is a function of the error terms; the moving average model links the current values of the time series to random errors that have occurred in the previous periods rather than the values of the actual series themselves. The moving average model can be written as:

$$y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Where  $\mu$  is the mean about which the series fluctuate,  $\theta'_s$  are the moving average parameters to be estimated, and  $\varepsilon_{t-q}$  's are the error terms (q=1, 2, 3,.....) assumed to be independently distributed over the time.

**iii. Mixed autoregressive and moving average model**

Under the assumption of stationary, the mixed autoregressive and moving average model of Box-Jenkins methodology is known ARMA model. In other words, the series is assumed stationary (no need for differencing) and the model is written as:

respectively, the model is referred to as ARMA (p, q).

#### iv. Mixed autoregressive, integrated and moving average (ARIMA) model

When the data is not stationary, then the Box-Jenkins (ARIMA) methodology is represented as ARIMA (p, d, q), where d denotes the degree of differencing involved to achieve stationarity in the series.

#### Time series regression model

The time series regression models relate the variable  $y_t$  to a function of time. Time series  $y_t$  can be described using a trend model; such model is defined as follow:

$$y_t = TR + \varepsilon_t = \beta_0 + \beta_1 t + \varepsilon_t$$

Where,  $y_t$ : The value of the time series in period t;  $TR$ : The trend in the time period t;  $\varepsilon_t$ : The error term in the time period.

Any one value of the error term is statistically independent of any other value of the error. In other words, serial correlation problem must not exist. To correct the serial correlation problem, Cochrane-Orcutt procedure is applied to produce better estimates. F-test statistic is used to test the overall fitness of the model while t-test is used to identify which variables to be included in the final model. Adjusted R-squared is then used to measure the goodness of fit of the model.

#### Models evaluation

The accuracy of model's performance is measured by the size of the forecast error. The operational meaning of an error is defined as the difference between the actual value and the fitted value generated from a given model by  $e = y - \hat{y}$ . So, the best method will give us the smallest error value.

#### Error measures

This study uses four measures of error that are commonly discussed and utilized by researchers and practitioners, namely, mean squared error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE) and geometric root mean squared error (GRMSE) (Armstrong and Collopy, 1992; Lazim, 2005).

#### Mean squared error (MSE)

$$MSE = \frac{\sum e^2}{n}$$

For which  $e = y - \hat{y}$  where  $y$  is the actual observation and  $\hat{y}$  is the estimated value. Its strength lies in its mathematical simplicity, that is, it is easy to understand and to calculate, it has the tendency to penalize large forecast errors more severely than other common accuracy measures to determine which method avoid large errors.

In other words, an incidence of a large error would significantly influence the value of MSE. But the main disadvantage the MSE faces is that it is easily affected by extreme values.

#### Root mean square error (RMSE)

This is the most favored measure among the practitioners and has even stronger preference among the academics.

$$RMSE = \sqrt{\frac{\sum e^2}{n}}$$

The RMSE gives equal weights to all errors; this can also be disadvantageous to a model that has one large forecast error. Thus, when an analyst ranks the forecasting models by RMSE, the presence of one or two extreme errors may alter the ranking of the model.

#### Mean absolute percentage error (MAPE)

$$MAPE = \sum_{t=1}^n \frac{|(e_t / y_t) \times 100|}{n}$$

The main disadvantage of this measure lies in its relevancy as it is valid only for the ratio-scaled data (that data with a meaningful zero). For this reason, MAPE is potentially explosive for large forecast error when the actual observations are close to zero.

#### Geometric root mean squared error (GRMSE)

$$GRMSE = \left( \prod_{t=1}^n e^2 \right)^{\frac{1}{2n}}$$

The existence of an outlier greatly affects the accuracy of the error measure. But the geometric root mean square error is the most useful alternative in this case.

#### Data collection

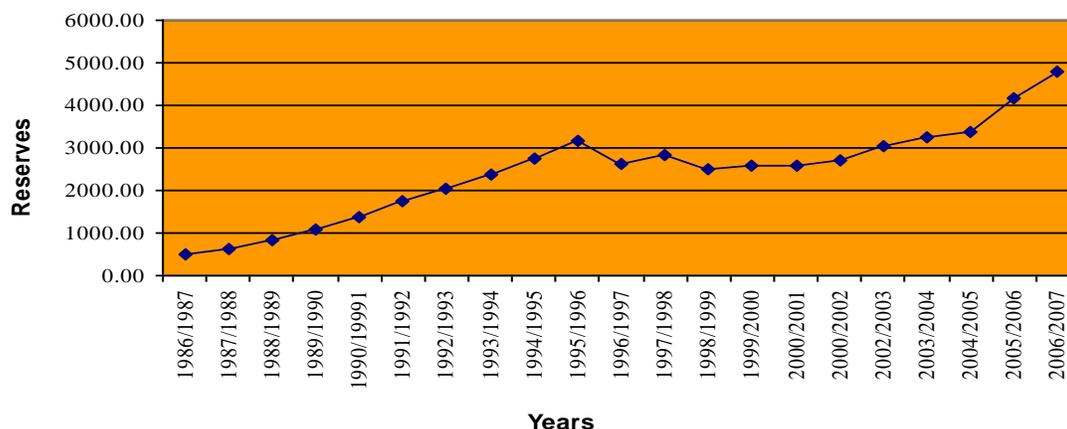
The Egyptian market is the object of this study, and the study will concentrate on loss reserves in general insurance segment within the years 1986/1987 till 2006/2007. Data encompasses of loss reserve value for the total general insurance industry. A total of 21 annual data is used.

#### FINDINGS

Figure 1 shows a plot of the yearly observations for loss reserves in the Egyptian market from the year 1986/1987 to the year 2006/2007. This plot exhibits a trend pattern because it appears to be an upward trend in the data. There is also rises and falls in this time series within the period 1995/1996 to 1998/1999 after that it continued to rise steadily until 2006/2007.

#### Models estimation

Out of 21 annual observations of loss reserves in the Egyptian market, 16 observations will be used to estimate the following forecasting models:



**Figure 1.** Loss reserves in general insurance in the Egyptian market. The vertical axis measures the variable loss reserves in the Egyptian market (dependent). The horizontal axis corresponds the time periods (independent). E.P, 000.

**Table 1.** Fitted model based on exponential smoothing technique.

T	Y	$\hat{y}$ 1- step	$\hat{y}$ 2- step	$\hat{y}$ 3- step	e 1- step	e 2- step	e 3- step
16 - 2002/2003	3054.06	2869.38240*			184.6776		
17- 2003/2004	3231.64	3214.03663**	3017.81880*		17.60337	213.8212	
18- 2004/2005	3392.11	3392.65412***	3374.01525**	3166.25520*	-0.54412	18.09475	225.8548
19- 2005/2006	4148.10	3553.09283****	3553.66824***	3533.99388**	595.0072	594.4318	614.1061
20- 2006/2007	4791.34	4340.40211*****	3714.07667****	3714.68235***	450.9379	1077.263	1076.658

\*1 – 16 observations used in the estimation process; \*\*1 – 17 observations used in the estimation process; \*\*\*1 – 18 observations used in the estimation process; \*\*\*\*1 – 19 observations used in the estimation process; \*\*\*\*\*1- 20 observations used in the estimation process.

1. Exponential smoothing model
2. Box-Jenkins (ARIMA) model
3. Time series regression model

These models forecasts will be compared to the actual values of the data, and the forecasting models will be ranked and compared depending on the standard error.

### Exponential smoothing technique

Based on the Holt method, the best “Alpha, Gamma and Delta” combination are: A = 1.00 alpha, G = 0.00 gamma. The fitted model is presented in Table 1.

### BOX-Jenkins model (ARIMA)

**Autocorrelation function (ACF) inspection:** As shown in Figure 2, the first spike of ACF values is large and it declines speedily to zero, therefore, we can conclude that the series is stationary, and it does not need to do

differencing, and the number of significant spikes equal 4, so MA(4) can represent this data series.

### Partial autocorrelation function (PACF) inspection:

Figure 3 shows there is significantly large spike followed by smaller other spikes at lags higher than 1. This means that the series is stationary with one significant spike, so, the AR (1) model can represent this data series.

**Fitting the ARIMA model:** For comparison purposes, 5 models were estimated, that is, ARIMA (1, 0, 4), ARIMA (1, 0, 3), ARIMA (1, 0, 2), ARIMA (1, 0, 1), and ARIMA (1, 0, 0). To determine which of the models fits the best, two criteria will be used; the first criteria is the value of AIC/SBC (Akaike’s information criteria) and the second, the value of the MSE. The results are summarized in Tables 2 and 3.

Hence it is concluded that ARIMA (1, 0, 2) model is relatively the best model since it has the smallest AIC and SBC. T- tests executed on ARIMA (1,0,2) model indicate that AR(1) can represent the model at 1%

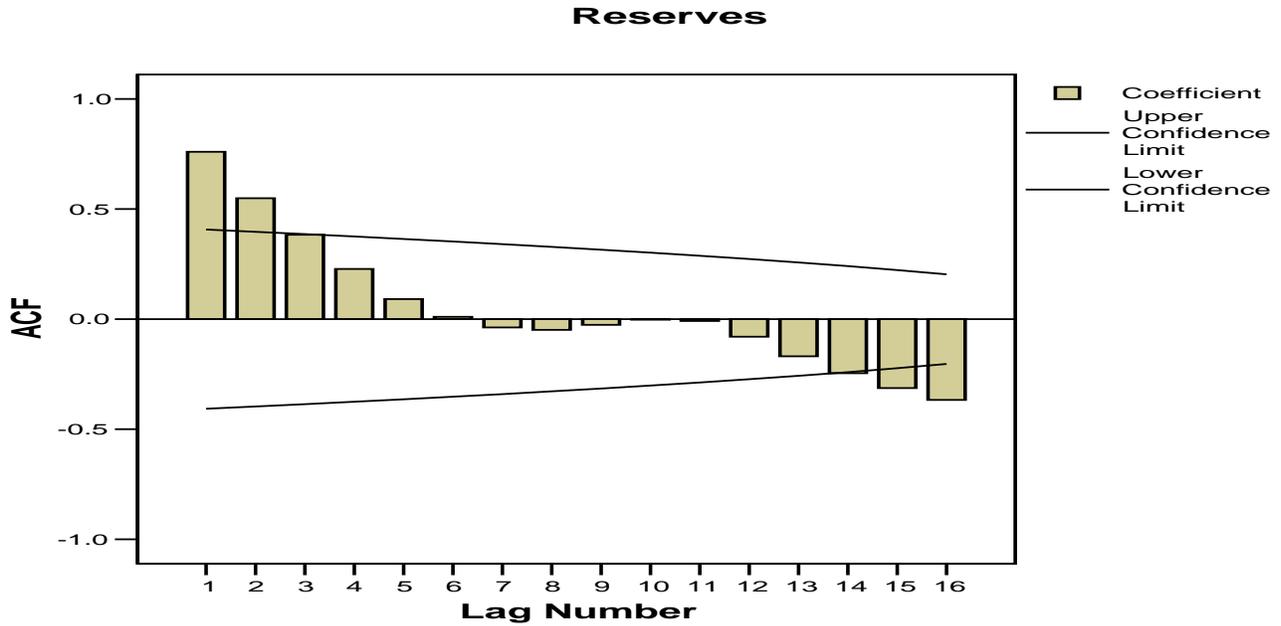


Figure 2. Autocorrelation function (ACF).

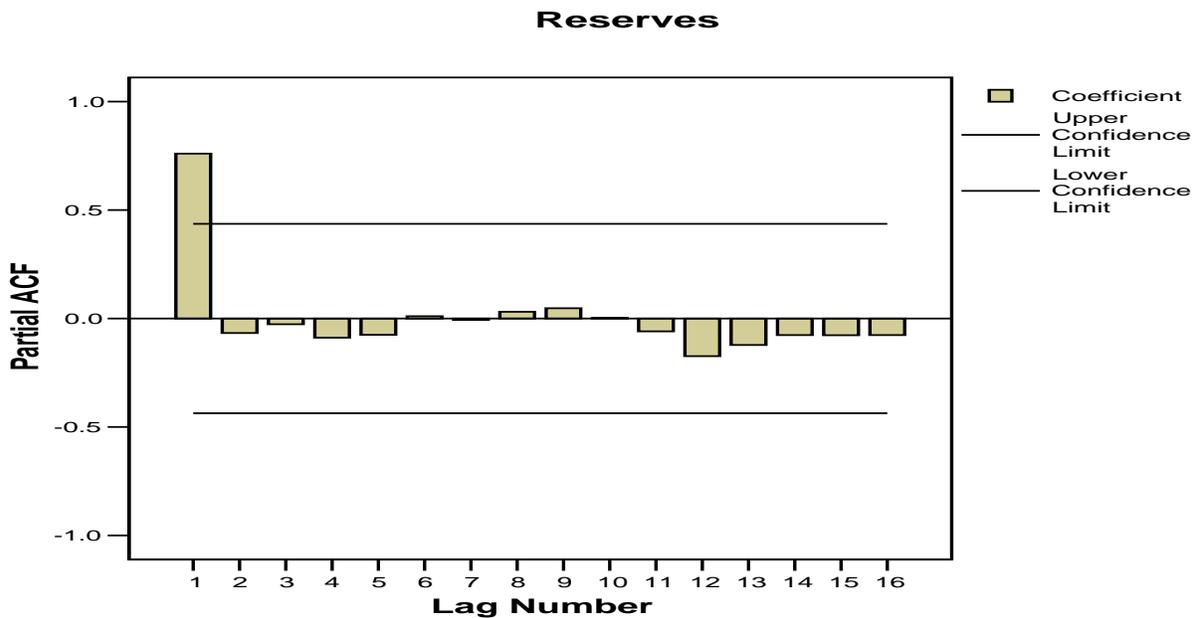


Figure 3. Partial autocorrelation function (PACF).

significant level (Table 4). Applying AR (1) forecasting model on the actual loss reserve data resulted in the estimated data ( $\hat{y}$ ) in Table 4.

**Time series regression**

Since the data series have a trend and does not have

seasonality (yearly data), the model that better represents the data is:  $y_t = TR + \epsilon_t$ . The linear model that represents the data series is therefore  $y_t = \beta_0 + \beta_1 t$

Normality analysis as presented in Table 5 and Figure 4 indicates that the distribution of loss reserve is close to a normal distribution. After correcting for serial correlation problem using Cochrane-Orcutt method, the final estimation model is presented in Table 6.

In summary, based on the Table 6, the following

**Table 2.** AIC and SBC summary table.

Statistic	Model form				
	ARIMA (1, 0, 4)	ARIMA (1, 0, 3)	ARIMA (1, 0, 2)	ARIMA (1, 0, 1)	ARIMA (1, 0, 0)
AIC	311.21441	311.8634	310.8158	313.04841	313.3809
SBC	317.48154	317.0861	314.9939	316.18198	315.4699

**Table 3.** Variables in the model.

Variable	B	SEB	T-RATIO	P-value
AR1*	0.96770	0.05960	16.237891	0.00000000
MA1	-0.35114	0.22226	-1.579884	0.13255884
MA2	-0.49334	0.25869	-1.907072	0.07355461
Constant	2795.43339	2054.26396	1.360796	0.19134728

\*represent significant at 1% level.

**Table 4.** Fitted model based on BOX-Jenkins model.

T	y	$\hat{y}$ 1- step	$\hat{y}$ 2- step	$\hat{y}$ 3- step	e 1- step	e 2- step	e 3- step
2002/2003	3054.06	2681.22528*			372.8347		
2003/2004	3231.64	3072.94776**	2691.55819*		158.6922	540.0818	
2004/2005	3392.11	3356.44472***	3175.45289**	2625.85818*	35.66528	216.6571	766.2518
2005/2006	4148.10	3396.71915****	3354.64058***	3107.67205**	751.3809	793.4594	1040.428
2006/2007	4791.34	3974.23214*****	3348.06791****	3287.10308***	817.1079	1443.272	1504.237

\*1 – 16 observations used in the estimation process; \*\*1 – 17 observations used in the estimation process; \*\*\*1 – 18 observations used in the estimation process; \*\*\*\*1 – 19 observations used in the estimation process; \*\*\*\*\*1- 20 observations used in the estimation process.  $y$  represents the actual loss reserve evaluation values.  $\hat{y}$  represents the forecasted value.  $e$  is the prediction error, calculated by taking the different between  $y$  and  $\hat{y}$ .

**Table 5.** Normality through SPSS.

Variable	Skewness			Kurtosis			Test of normality Kilmogorov-Smirnov		Description of distribution
	Statistic	Std. error	z-score (<2)	Statistic	Std. error	z-score (<7)	Statistics	Significance 0.05	
Reserves	-0.025	0.501	-0.05	-0.042	0.972	0.043	0.15	0.2	Normal

Z-stat skewness = skewness/std error skewness= skewness statistics/.501; Z-stat kurtosis = kurtosis/std error kurtosis = kurtosis statist.

equations will be used to predict y-values based on the series lengths:

$$y_t = \beta_0 + \beta_1 t$$

$$y_t = 2419.9237 + 32.2906 t \quad (\text{Series length: } 16)$$

$$y_t = 1711.6074 + 82.6809 t \quad (\text{Series length: } 17)$$

$$y_t = 1579.6242 + 92.9955 t \quad (\text{Series length: } 18)$$

$$y_t = 1497.4349 + 99.4705 t \quad (\text{Series length: } 19)$$

$$y_t = 885.05623 + 149.10163 t \quad (\text{Series length: } 20)$$

### Models evaluation and comparison

Each of the forecasting models is evaluated based on several error measures. The errors are reported in Tables 6 to 10.

A comparison is also made across results from different

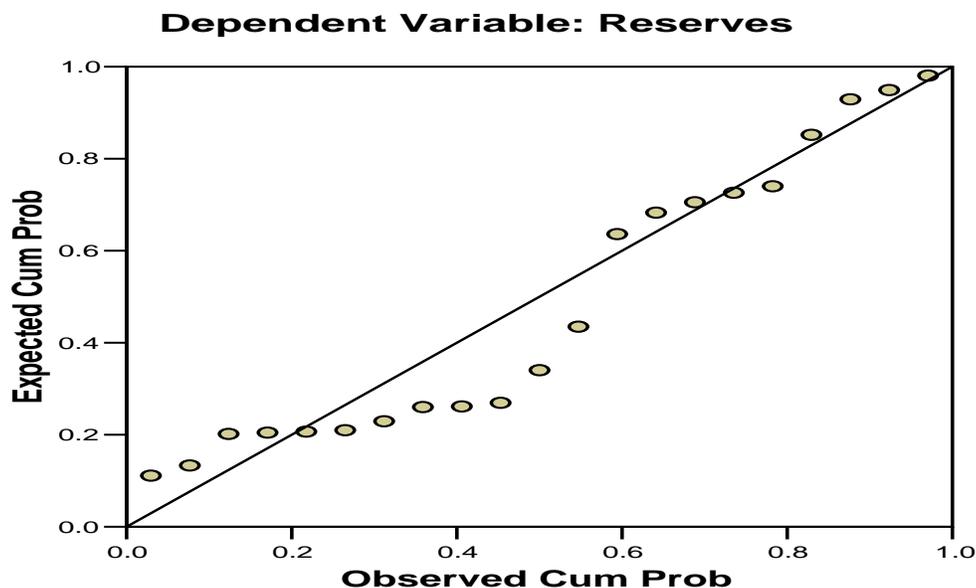


Figure 4. Normal p-p plot regression standardized residual.

Table 6. Time series regression models of loss reserve.

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		$\beta$	Std. Error	Beta		
16 Obs.	(Constant)	2419.9237	1382.9725		1.7498	0.106
	t	32.2906	94.0136	0.0987	0.34347	0.737
17 Obs.	(Constant)	1711.6074	1019.2068	0.3073	1.679	0.1169
	t	82.6809	71.0188		1.164	0.256
18 Obs.	(Constant)	1597.6242	894.61837	0.3974	1.766	0.099
	t	92.9955	60.60863		1.534	0.147
19 Obs.	(Constant)	1497.4349	803.81304	0.4377	1.8629	0.0822
	t	99.4705	52.75645		1.885	0.0789
20 Obs.	(Constant)	888.05623	654.82788	0.6459	1.35159	0.1953
	t	149.10163	44.05619		3.38435	0.0038

Table 7. Fitting the model.

T	y	$\hat{y}$ 1- step	$\hat{y}$ 2- step	$\hat{y}$ 3- step	e 1- step	e 2- step	e 3- step
2002/2003	3054.06	2968.864*			85.196		
2003/2004	3231.64	3199.864**	3001.155*		31.776	230.485	
2004/2005	3392.11	3346.539***	3282.545**	3033.445*	45.571	109.565	358.665
2005/2006	4148.10	3486.845****	3439.534***	3365.225**	661.255	708.566	782.875
2006/2007	4791.34	4016.19*****	3586.315****	3532.53***	775.15	1205.025	1258.81

\*1 – 16 observations used in the estimation process; \*\*1 – 17 observations used in the estimation process; \*\*\*1 – 18 observations used in the estimation process; \*\*\*\*1 – 19 observations used in the estimation process; \*\*\*\*\*1- 20 observations used in the estimation process.

**Table 8.** Exponential smoothing technique.

T	y	$\hat{y}$ 1- step	$\hat{y}$ 2- step	$\hat{y}$ 3- step	e 1- step	e 2- step	e 3- step
2002/2003	3054.06	2869.38240			184.6776		
2003/2004	3231.64	3214.03663	3017.81880		17.60337	213.8212	
2004/2005	3392.11	3392.65412	3374.01525	3166.25520	-0.54412	18.09475	225.8548
2005/2006	4148.10	3553.09283	3553.66824	3533.99388	595.0072	594.4318	614.1061
2006/2007	4791.34	4340.40211	3714.07667	3714.68235	450.9379	1077.263	1076.658
MSE					118358.9	389972.9	529109.7
RMSE					344.0333	624.4781	727.3993
GRMSE					54.35866	223.1037	530.5384
MAPE					6.072665	10.99092	14.64456

**Table 9.** BOX-Jenkins model (ARIMA).

T	y	$\hat{y}$ 1- step	$\hat{y}$ 2- step	$\hat{y}$ 3- step	e 1- step	e 2- step	e 3- step
2002/2003	3054.06	2681.22528			372.8347		
2003/2004	3231.64	3072.94776	2691.55819		158.6922	540.0818	
2004/2005	3392.11	3356.44472	3175.45289	2625.85818	35.66528	216.6571	766.2518
2005/2006	4148.10	3396.71915	3354.64058	3107.67205	751.3809	793.4594	1040.428
2006/2007	4791.34	3974.23214	3348.06791	3287.10308	817.1079	1443.272	1504.237
MSE					279539.9	762810.1	1310787
RMSE					528.7153	873.39	1144.896
GRMSE					264.54	605.0292	1062.429
MAPE					10.66751	18.08754	26.3554

**Table 10.** Time series regression.

T	y	$\hat{y}$ 1- step	$\hat{y}$ 2- step	$\hat{y}$ 3- step	e 1- step	e 2- step	e 3- step
2002/2003	3054.06	2968.864			85.196		
2003/2004	3231.64	3199.864	3001.155		31.776	230.485	
2004/2005	3392.11	3346.539	3282.545	3033.445	45.571	109.565	358.665
2005/2006	4148.10	3486.845	3439.534	3365.225	661.255	708.566	782.875
2006/2007	4791.34	4016.19	3586.315	3532.53	775.15	1205.025	1258.81
MSE					209692.1	504819.7	775378.8
RMSE					457.9215	710.5067	880.556
GRMSE					144.608	383.1975	707.0452
MAPE					7.447123	13.14847	18.57307

methods for one-step ahead; two-step ahead and three-step ahead. This evaluation to identify which method has

a better prediction. This evaluation will be presented in the Tables 11a, b and c.

**Table 11a.** One-step-ahead evaluation.

Technique	MSE	RMSE	GRMSE	MAPE	Total
Exponential	118358.9 (1)	344.0333 (1)	54.35866 (1)	6.072665 (1)	4
ARIMA	279539.9 (3)	528.7153 (3)	264.54 (3)	10.66751 (3)	12
TSR	209692.1 (2)	457.9215 (2)	144.608 (2)	7.447123 (2)	8

**Table 11b.** Two-step-ahead evaluation.

Technique	MSE	RMSE	GRMSE	MAPE	Total
Exponential	389972.9 (1)	624.4781 (1)	223.1037 (1)	10.99092 (1)	4
ARIMA	762810.1 (3)	873.39 (3)	605.0292 (3)	18.08754 (3)	12
TSR	504819.7 (2)	710.5067 (2)	383.1975 (2)	13.14847 (2)	8

**Table 11c.** Three-step-ahead evaluation.

Technique	MSE	RMSE	GRMSE	MAPE	Total
Exponential	529109.7 (1)	727.3993 (1)	530.5384 (1)	14.64456 (1)	4
ARIMA	1310787 (3)	1144.896 (3)	1062.429 (3)	26.3554 (3)	12
TSR	775378.8 (2)	880.556 (2)	707.0452 (2)	18.57307 (2)	8

From Table 11a, it is easy to say that the best forecasting technique for 1-step ahead is the exponential smoothing technique. Tables 11b and c shows that the best forecasting technique for 2-step and 3-step ahead is also the exponential smoothing technique.

These comparisons conclude that exponential smoothing technique is the best time series forecasting model for general insurance loss reserve in Egypt.

## Conclusion

The aim of this study is to apply time series analysis to forecast the loss reserves in general insurance in the Egyptian market. The forecasted loss reserves are one of the most important indicators that have many important and strategic decisions applications. For all the adopted models, the estimations are done by using the loss reserves data in general insurance in the Egyptian market for the period 1986 to 2006. The series from 1986 to 2001 were used for the estimations process and the other remaining observations (data from 2002 till 2006) are used to evaluate the models as outside sample data. The models are evaluated based on their accuracy in predicting outside sample data. The study concludes that the best model is the exponential smoothing technique in all steps-ahead.

## REFERENCES

Armstrong JS, Collopy F (1992). Error measure for generalizing about forecasting methods: Empirical comparison. *Int. J. Forecast.*, 8: 69-90.

- Bai Y, Barger C, Olandese A, Singh N, Sunesara R (2005). Estimating loss reserves for dental claims. Michigan State University, from: [www.msu.edu/user/baragar1/DeltaDental.doc](http://www.msu.edu/user/baragar1/DeltaDental.doc).
- Brown M (1996). Modeling and forecasting in insurance management. A guide to insurance management. edited by: Stephen Diacon. Macmillan, pp. 65-66.
- Bornhuetter RL, Ferguson RE (1972). The actuary and IBNR. *Proc. Casualty Actuarial Soc.*, 59: 181-195.
- Calandro J, O'Brien T (2004). A user-friendly introduction to property-casualty claim reserves. *Risk Manage. Insur. Rev.*, 7 (2): 177-187.
- Chan JSK, Choy STB, Makov UE (2004). The growing triangle technique for choosing a method of loss reserves. Proceedings of the 3rd Conference in Actuarial Science and Finance, held in Samos. on September 2-5. From: <http://www.stat.ucl.ac.be/Samos2004/mainpage.htm>
- Cheung D (1997). Estimating IBNR Reserves with Robust statistics. Unpublished Ph.D. Thesis. Western Michigan University, U.S.A.
- Douglas D (1999). Comparative Analysis of neural networks and traditional Actuarial methods for estimating casualty reserve liability. Unpublished Ph.D. Thesis. The University of Texas, U.S.A.
- England PD, Verrall RJ (2002). Stochastic claims reserving in general insurance. London: Institute of Actuaries, pp. 1-76.
- Grace MF, Laverty JT (2006). Property-Liability Insurer Reserve Error: Manipulation, Mistake, or Motive. World Risk and Insurance Economics Congress.
- Harbey G (1995). Using regression model technique in estimating outstanding claim reserve in property and liability insurance. *J. Financ. Commercial stud.*, Faculty of commerce. Beni Sueif. Cairo Univ., 5(1).
- Harnek RF (1966). Formula loss reserves. *Insurance Accounting and Statistical Proceedings*.
- Lazim MA (2005). *Introductory Business Forecasting*. Shah Alam: Uitm.
- Taylor GC (1977). Separation of inflation and other effects from the distribution of non-life insurance claim delays. *ASTIN Bull.*, 9: 219-230.