A Multivariate Model for Predicting the Efficiency of Financial Performance for Property and Liability Egyptian Insurance Companies

Osama Hanafy Mahmoud

Abstract: This paper uses the financial data of some property-liability insurance companies in Egypt to develop a multivariate model that reflects the efficiency of financial performance. Data will be classified statistically among three categories of financial performance based on the results of fuzzy clustering. The predictive variables in the multivariate model are represented as 25 financial ratios, which are more commonly used in describing the financial performance of the insurance company. Factor analysis has been utilized as a data reduction technique. The paper also examines the effect of the insurance company's ownership type (public vs. private). A multivariate model should be better able to identify the efficiency of financial performance by comparing the results of discriminant analysis and logistic regression.

Key Words: Financial performance, factor analysis, fuzzy clustering, discriminant analysis, logistic regression, property and liability insurance, Egyptian insurance companies

1. INTRODUCTION

Insurance companies sell protection to policyholders against many types of risks: property damage or loss, health and casualty, financial losses, etc. In return for this risk protection, insurance companies receive a premium from the policyholder that is used to cover expenses and the expected risk. For longer-term risk protections, part of the premiums is invested to get higher yields. Although the protection buyer mitigates the individual risk to the large and better-diversified portfolio of the insurer, the risk is not completely reduced because the insurer may default his obligations. Insurers need to have sufficient equity or buffer capital to meet their obligations in adverse conditions when their losses on the diversified portfolio exceed the expected losses. Ratings provide an assessment of the ability of the insurer to meet its obligations to policyholders and debt holders.

This study presents a model for identifying the financial performance of insurance companies. In this paper, financial ratios are used to describe and predict the financial performance of insurers. Beaver (1966) uses the financial ratios as predictors of failure and states that the usefulness of ratios can only be tested with regard to some particular purpose. Ratios are currently in widespread use as predictors of failure. While this is not the only possible use of ratios, it is a starting point from which to build an empirical verification of ratio analysis. Van Gestel et al. (2007) analyze the relationship between financial ratios and the rating for different types of insurance companies by using advanced statistical techniques that are able to detect non-linear relationship.

Various statistical models for the classification and prediction of financial performance have been presented in prior studies. Harrington and Nelson (1986) used regression analysis to estimate the relationship between premium-to-surplus ratios and insurer characteristics, including asset and product mix variables. Analysis of the regression residuals then can be used to identify insurers with ratios that are substantially higher than those for insurers with similar characteristics. The method is illustrated using data for solvent and insolvent insurers. The method's ability to identify insurers that later became insolvent is compared to that of the National Association of Insurance Commissioners (NAIC) Insurance Regulatory Information System. BarNiv and Hershbarger (1990) presented models that incorporate variables designed to identify the financial solvency of life insurers. Three multivariate analyses (multidiscriminant, nonparametric, and logit) have been used to examine the applicability and efficiency of alternative multivariate models for solvency surveillance of life insurers. Thus, decision tools are developed based on real data, and systematic statistical frameworks are applied for evaluating the financial viability of life insurers. A comprehensive review of insurer insolvency literature is given in BarNiv and McDonald (1992).

Ambrose and Carroll (1994) examined the efficiency of Best's recommendations, Insurance Regulatory Information System (IRIS) ratios, and other financial measures for their statistical ability to classify solvent and insolvent life insurers. Ambrose and Carroll estimated classification models for a sample of insurers for 1969 through 1986 and applied the models to a holdout sample for 1987 through 1991. The financial variables and IRIS ratios outperformed Best's recommendations in distinguishing between the two groups in a logit model. Lee and Urrutia (1996) compared the performance of the logit and hazard models in predicting insolvency and detecting variables that have a statistically significant impact on the solvency of property-liability insurers. The empirical results indicated that the hazard model identifies more significant variables than the logit model and that both models have comparable forecasting accuracy.

Using estimations across 18 lines of insurance for the years 1984 through 1993, Chidambaran et al. (1997) presented an empirical analysis of the economic performance of the U.S. property-liability insurance industry. They adopted an industrial organization approach, focusing on the economic loss ratio as a measure of pricing performance. The concentration ratio for the line and the share of direct writers in the line were both found to be significant determinants of performance. The results were consistent with shortcomings in competition in some insurance lines. On a sample of forty-eight insolvent life insurers over the period 1990 to 1992, Pottier (1998) compared the predictive abilities (1) ratings, rating changes and total assets; (2) financial ratios; and (3) financial ratios combined with ratings and rating changes. Based on the expected cost of misclassification, the predictive abilities of ratings, rating changes, and total assets are comparable to financial ratios combined with ratings and rating changes. For most cost ratios, combining ratings and rating changes with financial ratios improved predictive ability compared to using financial ratios alone. Another interesting finding is that adverse rating changes are important predictors of insolvency.

BarNiv et al. (1999) illustrated a method that constructs confidence intervals for insolvency probabilities and examined various measures of the confidence intervals, such as their minimum lengths and minimum upper bounds. Lai and Limpaphayom (2003) examined the impact of organizational structure on firm performance, incentive problems, and financial decisions in the Japanese nonlife (property-casualty) insurance industry.

The remainder of this paper is organized as follows:

An introduction for clustering is presented in section 2. The methodology for clustering and fuzzy classification is presented in section 3. The proposed multivariate models for financial performance (discriminant analysis model and logistic regression model) are presented in sections 4 and 5, respectively. Section 6 provides the application that develops a multivariate model based on financial data of Egyptian property-liability insurance companies. This application reflects the efficiency of financial performance.

2. INTRODUCTION To CLUSTERING

Cormack (1971) presented a review of classification by proposing the definitions of similarity and of cluster. The principles, but no details of implementation, of the many empirical classification techniques currently in use are discussed. Limitations and short comings in their development and practice are also pointed out.

Clustering is the classification of objects into different groups, or more precisely, the partitioning of a data set into subsets (clusters) so that the data in each subset (ideally) share some common trait, often proximity according to some defined distance measure. Data clustering is a common technique for statistical data analysis that is used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics [51]. Therefore, the cluster is a group of cases representing an access intensity within the population at risk that is unlikely to be due to chance (Alexander (1999)).

Cluster analysis is an important self-study tool for solving the problem of finding groups in data without the help of a response variable (Tibshirani et al. (2001)).

The goal of clustering or classification is to decide which of two or more populations a particular observation or set of observations belongs. In this type of problem, an error is made if an observation is assigned to any population other than the one to which it belongs. Thus, the relative value of a classification or assignment procedure can be measured in two ways: (1) assume that the underlying distributions are known and compute the probability of assigning an observation to the wrong population, or (2) use the procedure on a set of observations for which the correct assignments are known and calculate the percentage of observations that are assigned to a wrong population (Mayer (1971)).

2.1 Common Distance Functions

An important step in any clustering is to select a distance measure that will determine how the similarity of two elements is calculated. This will influence the shape of the clusters, as some elements may be close to one another according to one distance and further away according to another.

2.1.1 Euclidean distance

Euclidean distance is the straight line distance between two points. In a plane with p_1 at (x_1, y_1) and p_2 at (x_2, y_2), it is

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

The Euclidean distance function measures the distance between a point $\boldsymbol{X}$ ($\boldsymbol{X_1}$, $\boldsymbol{X_2}$, etc.) and a point $\boldsymbol{Y}$ ($\boldsymbol{Y_1}$, $\boldsymbol{Y_2}$, etc.) as

$$d = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2},$$

where n is the number of variables, and $\boldsymbol{X_i}$ and $\boldsymbol{Y_i}$ are the values of the ith variable, at points $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively.

2.1.2 Manhattan distance

Manhattan distance is the distance between two points measured along axes at right angles. In a plane with p_1 at (x_1, y_1) and p_2 at (x_2, y_2), it is $|x_1 - x_2| + |y_1 - y_2|$.

The Manhattan distance function computes the distance that would be traveled to get from one data point to the other if a grid-like path is followed. The Manhattan distance between two items is the sum of the differences of their corresponding components.

The formula for this distance between a point $\boldsymbol{X}$=($\boldsymbol{X_1}$, $\boldsymbol{X_2}$, etc.) and a point $\boldsymbol{Y}$=($\boldsymbol{Y_1}$, $\boldsymbol{Y_2}$, etc.) is:

$$d = \sqrt{\sum_{i=1}^{n} |x_i - y_i|}$$

2.1.3 Mahalanobis distance

Mahalanobis distance is most commonly used as a multivariate outlier statistic. This measure is recommended for examining data profiles such as learning curves, serial position effects, and group profiles.

The Mahalanobis distance has the advantage of utilizing group means and variances for each variable, and the correlations and covariance between measures. The Mahalanobis distance matrix algebra equation is written as follows:

$$\left(X_i - Y_i\right)' S_i^{-1}\left(X_i - Y_i\right)$$

where S^{-1} is the inverse covariance matrix.

Estimating the number of clusters represent the main problem when using the cluster analysis. Authors interested by this problem include Tibshirani et al. (2001) and Peck et al. (1989). Tibshirani et al. proposed a method (the "gap statistic") for estimating the number of clusters (groups) in a set of data. Peck et al. developed a bootstrap-based procedure for obtaining approximate confidence bounds on the number of clusters in the "best" clustering. The effectiveness of this procedure is evaluated in a simulation study. An application is presented.

2.2 Clustering Algorithms Used for Classification

2.2.1 *K*-Means Clustering

By assuming k clusters and defining k centroids, one for each cluster, this algorithm aims to minimize the objective function, which represents the squared error function in the following form:

$$S = \sum_{j=1}^{k}\sum_{i=1}^{n}\left\| x_i^{(j)} - C_j \right\|^2 .$$

Where $\left\| x_i^{(j)} - C_j \right\|$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster center C_j.

2.2.2 Hierarchical Clustering

Hierarchical algorithms find successive clusters using previously established clusters, whereas partitional algorithms determine all clusters at once. Hierarchical algorithms can be agglomerative ("bottom-up") or divisive ("top-down"). Agglomerative algorithms begin with each element as a separate cluster and merge them into successively larger clusters. Divisive algorithms begin with the whole set and proceed to divide it into successively smaller clusters.

2.2.3 Fuzzy Clustering

Clustering involves the task of dividing data points into homogeneous classes or clusters so that items in the same class are as similar as possible and items in different classes are as dissimilar as possible. In non-fuzzy or hard clustering, data is divided into crisp clusters, where each data point belongs to exactly one cluster. In fuzzy clustering, the data points can belong to more than one cluster. Associated with each of the points are membership grades that indicate the degree to which the data points belong to the different clusters.

Many authors in insurance use clustering. Some examples include Jensen (1971), who studied the financial performance of selected business firms using cluster analysis and Samson (1986), who depended on the classification system for designing an automobile insurance. Fiegenbaum and Thomas (1990) studied the performance of the U.S. insurance industry by using clustering to make the strategic groups. Yeo et al. (2001) used clustering technique for classifying risks and predicting claim costs in the automobile insurance industry. Wagstaff and Lindelow (2008) studied the effect of insurance on increased financial risk for health insurance in China.

3. FUZZY CLUSTERING

There are many text books that present the concept of fuzzy clustering and its algorithms and applications, such as Oliveira and Pedrycz (2007), Lazzerini et al. (2000), Pedrycz (2005), Sato et al. (1997), and the *NCSS Manual* (2007).

Fuzzy clustering generalizes partition clustering methods by allowing an individual to be partially classified into more than one cluster. In regular clustering, each individual is a member of only one cluster. Suppose we have *K* clusters and we define a set of variables, $m_{i1}, m_{i2}, \ldots, m_{ik}$, that represent the probability that object *i* is classified into cluster *k*. In partition clustering algorithms, one of these values will be one and the rest will be zero. This represents the fact that these algorithms classify an individual into one and only one cluster (Kaufman and Rousseeuw (1990)).

In fuzzy clustering the membership is spread among all clusters. The m_{ik} can now be between zero and one, with the stipulation that the sum of their values is one. We call this a fuzzification of the cluster configuration. It has the advantage that it does not force every object into a specific cluster.

The fuzzy algorithm seeks to minimize the following objective function, *C*, made up of cluster memberships and distances.

$$C = \sum_{L=1}^{k} \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} m_{iL}^2 m_{jL}^2 d_{ij}}{2\sum_{j=1}^{N} m_{jL}^2}$$

where m_{iL} represents the unknown membership of the object *i* in cluster *L* and d_{ij} is the dissimilarity between objects *i* and *j*.

The memberships are subject to constraints that they all must be non-negative and that the memberships for a single individual must sum to one. That is, the memberships have the same constraints that they would if they were the probabilities that an individual belongs to each group.

To test goodness of fit for fuzzy clustering, Kaufman and Rousseeuw (1990) proposed the silhouette statistic for assessing clusters and estimating the optimal number. For observation *i*, let *a*(*i*) be the average distance to other points in its cluster, and b(i) the average distance to points in the nearest cluster besides its own nearest is defined by the cluster minimizing this average distance. Then the silhouette statistic is defined by

$$S(i) = \frac{b(i) - a(i)}{Max \quad \{a(i), \; b(i)\}}$$

A point is well clustered if *s*(*i*) is large. Kaufman and Rousseeuw (1990) proposed to choose the optimal number of clusters $\hat{k}$ as the value maximizing the average *s*(*i*) over the data set.

Various studies used the fuzzy clustering in insurance such as Lemaire (1990), who indicated the role of the fuzzy theory in decision making for calculating net premiums and reinsurance polices. Ebanks et al. (1992) presented how to use the measures of fuzziness to risk classification for life insurance in two phases—the first determinate is the degree of risk and the second determinate is the membership for each risk. Cummins and Derrig (1993) used the fuzzy trends in estimating the property-liability insurance claim costs. Carreno and Jani (1993) developed expert system based on fuzzy approach to insurance risk assessment.

Young (1996 and 1997) introduced the fuzzy approach for adjusting and changing the insurance rates. Verrall and Yakoubov (1998) used fuzzy clustering to grouping by policyholder age in general insurance. The fuzzy approach is used for estimating the reserves by Sanchez and Gomez (2003). Shapiro (2004) presented an overview of insurance studies that use fuzzy logic in numerous of applications for classification, underwriting, projected liabilities, fuzzy future and present values, pricing, asset allocations and cash flows, and investments. Sánchez (2006) and (2007) used the fuzzy regression to calculate insurance claim reserves.

4. DISCRIMINANT ANALYSIS

The statistical pattern recognition for discriminant analysis and how to apply it proposed by various studies as the *NCSS Manual* (2007), Huberty and Olejnik (2006), McLachlan (2004) and Huberty (1994).

Discriminant analysis finds a set of prediction equations based on independent variables used to classify individuals into groups. There are two possible objectives in a discriminant analysis: finding a predictive equation for classifying new individuals or interpreting the predictive equation to better understand the relationships that may exist among the variables.

In many ways, discriminant analysis parallels multiple regression analysis. The main difference between these two techniques is that regression analysis deals with a continuous dependent variable, while discriminant analysis must have a discrete dependent variable. The methodology used to complete a discriminant analysis is similar to regression analysis. You plot each independent variable versus the group variable. You often go through a variable selection phase to determine which independent variables are beneficial. You conduct a residual analysis to determine the accuracy of the discriminant equations.

Discriminant analysis assumes linear relations among the independent variables. Suppose you have data for K groups, with N_k observations per group. Let N represent the total number of observations. Each observation consists of the measurements of p variables. The i^{th} observation is represented by X_{ki}. Let M represent the vector of means of these variables across all groups and M_k the vector of means of observations in the k^{th} group.

Define three sums of squares and cross products matrices, S_T, S_W, and S_A, as follows:

$$S_T = \sum_{L=1}^{K} \sum_{i=1}^{N_L} (X_{Li} - M)(X_{Li} - M)^{\backslash}$$

$$S_W = \sum_{L=1}^{K} \sum_{i=1}^{N_L} (X_{Li} - M_L)(X_L i - M_L)^{\backslash}$$

$$S_A = S_T - S_W \;.$$

A discriminant function is a weighted average of the values of the independent variables. The weights are selected so that the resulting weighted average separates the observations into the groups. High values of the average come from one group; low values of the average come from another group. The problem reduces to one of finding the weights that, when applied to the data, best discriminate among groups according to some criterion. The solution reduces to finding the eigenvectors, V, of $S_W^{-1} S_A$. The canonical coefficients are the elements of these eigenvectors.

A goodness-of-fit parameter, Wilks' Lambda, is defined as follows:

$$\Lambda = \frac{|S_W|}{|S_T|} = \prod_{j=1}^{m} \frac{1}{1+\lambda_j}$$

where is the j^{th} eigenvalue corresponding to the eigenvector described above and m is the minimum of K-1 and p.

The canonical correlation between the j^{th} discriminant function and the independent variables is related to these eigenvalues as follows:

$$r_{ci} = \sqrt{\frac{\lambda_i}{1+\lambda_i}}$$

The linear discriminant functions are defined as:

$$LDF_k = W^{-1} M_k$$

Where

$$W = \left(\frac{1}{N-K} \right) S_W$$

Many studies in insurance have used discriminant analysis. For example, Trieschmann and Pinches (1973) used multiple discriminant analysis to classify firms into two groups (solvent or distress). The model developed was able to classify correctly forty-nine out of fifty-two firms included in the study. One solvent firm was classified as being in distress while two of the distressed firms were classified as belonging to the solvent group. The six variables used to classify firms were (1) agents balances/total asset ratio, (2) stock cost (preferred and common)/stock-market ratio (preferred and common), (3) bond cost/bonds-market ratio. (4) loss adjustment expenses paid -*k* underwriting expenses paid/net premiums written ratio, (5) combined ratio, and (6) premiums written direct/surplus ratio. Trieschmann and Pinches (1974) later examined the efficiency of alternative models for solvency surveillance of property-liability insurance firms employing financial ratios. The two models they investigated were (1) financial ratios individually or in groups on a univariate basis, and (2) a set of financial ratios in a multivariate context based on a multiple discriminant model. Through the use of statistical tests, it is shown that the multiple discriminant model does a better job of identifying firms with a high probability of distress than the univariate models. Ambrose and Seward (1988) incorporated Best's ratings into the discriminant analysis through a system of dummy variates. Best's ratings are then compared to the results obtained by the use of financial variables. Finally, a two-stage discriminant technique is introduced and its results are shown to be better for predicting insolvency for property-liability firms.

5. LOGISTIC REGRESSION

The application of logistic regression in the social sciences and its properties are proposed by many authors such as Kleinbaum et al. (2005), Menard (2001), Jaccard (2001), and Hosmer and Lemeshow (2000), as well as by the *NCSS Manual* (2007).

Logistic regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name logistic regression is often used when the dependent variable has only two values. The name multiple-group logistic regression (MGLR) is usually reserved for the case when the dependent variable has three or more unique values.

Logistic regression competes with discriminant analysis as a method for analyzing discrete response variables. In fact, the current feeling among many statisticians is that logistic regression is more versatile and better suited for most situations than is discriminant analysis because it does not assume that the independent variables are normally distributed, as discriminant analysis does.

In multiple-group logistic regression, a discrete dependent variable Y having G unique values ($G \geq 2$) is regressed on a set of p independent variables, $X_1, X_2, \ldots, X_p$. Y represents a way of partitioning the population of interest. For example, Y may be the condition of the financial performance for the insurance company.

The logistic regression model is given by the G equations as follows:

$$\log it\,(y = g \mid X) = X\beta$$

$$X = (X_1, X_2, \ldots, X_p)$$

where

$$\beta_g = \begin{bmatrix} \beta_{g1} \\ \vdots \\ \beta_{gp} \end{bmatrix}$$

and p_g is the probability that an individual with values $X_1, X_2, \ldots, X_p$ is in group g.

$$p_g = \frac{\exp(X\beta)}{1 + \exp(X\beta)}$$

We can use the maximum likelihood method to estimate the parameters for the logistic regression (β) as follows:

Let

$$\pi_{gj} = \text{prob}(Y = g \mid X_j) = \frac{\mathbf{exp}(X_j \beta_g)}{\sum_{s=1}^{G} \mathbf{exp}(X_j \beta_g)}.$$

The likelihood function is:

$$L = \prod_{j=1}^{N} \prod_{g=1}^{G} \pi_{gj}^{\ y_{gj}}$$

Where y_{gj} is one if the j^{th} observation is in group g and zero otherwise.

The log likelihood function is given by:

$$\ln L = \sum_{j=1}^{N} \sum_{g=1}^{G} y_{gj} Ln(\pi_{gj})$$

$$= \sum_{j=1}^{N} \left[\sum_{g=1}^{G} y_{gj} X_j \beta_g - \ln\left(\sum_{g=1}^{G} \exp(X_j \beta_g) \right) \right].$$

Maximum likelihood estimates of the βs are found by finding those values that maximize this log likelihood equation. This is accomplished by calculating the partial derivatives as:

$$\frac{\partial \ln L}{\partial \beta_{ik}} = \sum_{j=1}^{N} X_{kj} (y_{ig} - \pi_{ig})$$

$$g = 2,3,\ldots G \quad \text{and} \quad k = 1,2,\ldots,p$$

Setting these equations to zero solves it. Because of the nonlinear nature of the parameters, there is no closed-form solution to these equations and they must be solved iteratively by using Newton-Raphson method.

Several studies have used the logistic regression in the area of insurance. Beirlant et al. (1992) applied logistic regression to determine different subportfolios and adjusted insurance premiums for contracts belonging to a more or less heterogeneous portfolio, based on a representative sample from Belgian car insurance data from 1989. Devaney (1994) illustrated the usefulness of financial ratios as predictors of household insolvency by applying the logistic regression for forecasting. Ambrose and Carroll (1994) applied logistic regression analysis to matched-pair samples of life insurers and found that financial variables combined with NAIC Insurance Regulatory Information System ratios outperformed A.M. Best's recommendations in distinguishing between solvent and insolvent insurers. Combining all three types of predictors into one model provided the most accurate classification.

Steven and David (2000) analyzed the ability to forecast for NAIC and Best's company solvency measures by using logistic regression. Esteban and Jose (2001) used logistic regression for insurance risk classification. Mosley (2005) concentrated on the use of predictive modeling as logistic regression in the insurance industry. Cooper and Zheng (2007) and Zheng at al. (2007) used the logistic regression for estimating the contributions of speeding and impaired driving to insurance claim cost.

6. ApplicationS for Property-Liability Insurance Companies In Egypt

The initial data set consists of the financial ratios of property and liability insurance firms that represent the predictive variables for six insurance companies, three of which are public sector (Misr Insurance Company, Al-Chark Insurance Company and National Insurance Company). The others represent private sector companies (Suez Canal Insurance Company, El-Mohandes Insurance Company, and Delta Insurance Company). Data are derived from the annual statements of the period from 1992/1993 to 2005/2006.

6.1 Predictive Variables

Financial ratios provide a quick and relatively simple means of examining the financial condition of a business. A ratio simply expresses the relation of one figure appearing in the financial statements to some other figure appearing there or perhaps to some resource of the business. Ratios can be very helpful when comparing the financial health of different businesses. By calculating a relatively small number of ratios, it's often possible to build up a reasonably good picture of the position and performance of a business. There is no generally accepted list of ratios that can be applied to financial statements, nor is there a standard method of calculating many of them. In this paper, the predictive variables are the financial ratios for property and liability insurance firms, which are defined in the following table.

Variable	Definition
x_1	Net premiums written/surplus Measures the underwriting risk where a high ratio indicates greater underwriting exposure and more risk.
x_2	Adjusted surplus/surplus Measures the effects in the reinsurance policies and its reflection on insurer's retention rates.
x_3	Surplus/investment income Measures the ratio between the income from insurance operations and income from investments.
x_4	Net premiums written/premiums written Measures the retention policy of insurer.
x_5	Return on investments Measures the efficiency of insurer's investment portfolio.
x_6	Net profit/total assets Expresses the relationship between the net profit the business generates and total assets in the balance sheet.
x_7	Net profit/surplus Reflects the returns of an insurance company and the effect of its investments and underwriting policies.
x_8	Total liabilities/liquid assets Measures the ability of an insurer to have sufficient liquid recourses available to meet maturing obligations.
x_9	Debtors and sundry debtors/surplus Measures an insurer's ability to pay its policyholders' claims and to meet its other financial obligations.
x_{10}	Debtors and sundry debtors/total assets Describes the proportion of other people's money to the total claims against the assets of the business. The higher the ratio, the greater the likely risk for lender.
x_{11}	Reserves/surplus Measures the ability of an insurer to estimate reserves.
x_{12}	(Reserves + Surplus)/Net premiums written Measures the insurer's ability to cover the exposed money to risk and it's the underwriting policy.
x_{13}	Reserves/liquid assets Measures the ability of insurers to own assets that are easily converted to cash and to pay its liabilities on time.
x_{14}	Total liabilities/total assets Measures the portion of assets that are financed by others; its complement is the equity ratio.

Variable	Definition
x_{15}	Underwriting expenses paid/premiums written Measures an insurer's fund flow from insurance operations.
x_{16}	Commissions paid/premiums written Measures an insurer's fund flow from production operations.
x_{17}	Underwriting expenses paid total/premiums written Measures an insurer's fund flow from underwriting and production operations.
x_{18}	Loss ratio Measures the ratio between the premiums paid to an insurance company and the claims settled by the company.
x_{19}	Reinsurance ceded premiums/direct premiums written Measures an insurer's reinsurance policy and its acceptable appetite to cover different risks.
x_{20}	Cash/adjusted surplus This variable examines the ratio of adjusted surplus to convert to cash which help in planning for liquidity insurer's policy.
x_{21}	Total liabilities/adjusted surplus Measures the contribution of the insurer's adjusted surplus used to cover its total liabilities.
x_{22}	Net premiums written/total assets Examines the effectiveness of the assets a business employs in retaining premiums.
x_{23}	Direct premiums written/surplus Measures the underwriting risk for direct insurance operations.
x_{24}	Claims paid/liquid assets Compares liquid assets of a business with the current liabilities.
x_{25}	Claims paid/reserves Measures adequate the estimated reserves the meet claims paid.

Where

liquid assets = cash + deposits + reserves

Adjusted surplus = surplus – (Ceded reinsurance unearned premium × (Reinsurance ceded/ premiums written ceded)).

To test of the effect of an insurance company's ownership type (public vs. private) on the efficiency of financial performance, we need to know if the public and private sectors have the same strategies for developing their financial performance efficiencies or if their strategies are different. Thus, we will test the following hypotheses:

H_0: There is no difference between the mean of efficiency of financial performance ratios in the public sector and private sector.

H_1: There is a difference.

Table 1--The *t*-test for financial performance efficiency ratios in Egyptian insurance sectors

Ratios	*t*	Sig.
x_1	-4.360	0.000
x_2	6.245	0.000
x_3	-2.516	0.014
x_4	4.366	0.000
$\boldsymbol{x_5}$	**1.356**	**0.179**
$\boldsymbol{x_6}$	**-0.995**	**0.322**
$\boldsymbol{x_7}$	**0.732**	**0.466**
x_8	-3.149	0.002
x_9	-3.650	0.001
x_{10}	-6.572	0.000
x_{11}	5.465	0.000
x_{12}	12.043	0.000
x_{13}	5.331	0.000
$\boldsymbol{x_{14}}$	**-1.052**	**0.296**
$\boldsymbol{x_{15}}$	**-1.963**	**0.053**
x_{16}	-10.089	0.000
x_{17}	-6.923	0.000
x_{18}	2.036	0.045
x_{19}	-2.635	0.010
x_{20}	-3.893	0.000
x_{21}	-2.490	0.015
x_{22}	-6.319	0.000
x_{23}	-5.291	0.000
x_{24}	-3.259	0.002
x_{25}	-2.820	0.006

From Table 1, we can conclude that there is not a significance difference between the mean for efficiency of financial performance ratios in the public sector and the private sector for the following ratios:

Return on investments

Net profit/total assets

Net profit/surplus

Total liabilities/Total assets

Underwriting expenses paid/premiums written

6.2 Factor Analysis

The main applications of factor analytic techniques are to reduce the number of variables and detect structure in the relationships between variables, that is, to classify variables. Therefore, factor analysis is applied as a data reduction or structure detection method.

We used the statistical package SPSS to perform a factor analysis on efficiency and financial performance data, consisting of 25 ratios. Our results concluded that the data can be reduced to six factors when applying extraction method principle component analysis and Varimax with Kaiser Normalization rotation method, the most popular orthogonal technique.

Table 2 shows the rotated component and component score coefficient for each variable with related factor and its percent of variance.

Table 2—Rotated Component and Component Score Coefficient for Each Variable with Related Factors

Factor	% of Variance	Variable	Rotated Component	Component Score Coefficient
F1	19.430	x_{19}	0.934	0.293
		x_4	-0.857	-0.269
		x_2	-0.777	-0.164
		x_{21}	0.707	0.134
		x_{23}	0.662	0.099
		x_{20}	0.626	0.166
F2	18.568	x_{22}	0.931	0.286
		x_{24}	0.889	0.303
		x_{12}	-0.708	-0.136
		x_1	0.630	0.082
F3	18.335	x_{11}	0.899	0.269
		x_{13}	0.820	0.225
		x_3	-0.775	-0.199
		x_{14}	0.628	0.127
		x_8	0.572	0.123
		x_5	0.551	0.076
F4	14.268	x_{17}	0.891	0.301
		x_{16}	0.844	0.267
		x_{15}	0.661	0.248
		x_{10}	0.625	0.194
		x_9	0.566	0.244
F5	8.458	x_6	0.934	0.459
		x_7	0.882	0.427
		x_{18}	-0.562	-0.261
F6	4.741	x_{25}	-0.698	-0.583

We can estimate each factor's predictive value using regression analysis through the score coefficient of each variable belonging to its factor, which is a method we will use in the following section to cluster our cases according to their financial performance efficiencies.

6.3 Fuzzy Clustering

When distance is measured in Euclidean distance, we used NCSS (Number Cruncher Statistical Software) to determine our data's optimal number for fuzzy clustering. Table 3 shows our results.

Table 3—The Numbers of Fuzzy Clustering and Its Silhouette Statistic

Number of clusters	s(i)
2	0.161224
3	0.186804
4	0.049518
5	-1.00000

Table 3 shows that the value maximizing the average s(i) over the data set is 0.186804 at 3, which is the optimal number of clusters.

Therefore, we can organize our data into three clusters that can be classified according to its medoids cluster into three levels of efficiency for financial performance (low, moderate, and high), as shown in the Table 4.

Table 4—The Number and Percent of Cases According To Fuzzy Classification for Efficiency of Financial Performance

Efficiency of financial performance	No.	%
Low	29	34.5
Moderate	23	27.4
High	32	38.1
Total	**84**	**100**

Table 5 represents the cross-tabulation between the ownership type of the insurance company (public or private) and the fuzzy classification for efficiency of financial performance.

Table 5—Number and Percentage of Cases Based on Fuzzy Classification for Efficiency of Financial Performance and the Insurance Company's Ownership Type

Sector	Efficiency of Financial Performance						Total
	Low		Moderate		High		
	No.	%	No.	%	No.	%	
Public	28	66.7	2	4.8	12	28.6	42
Private	1	2.4	21	50	20	47.6	42
Total	**29**	**34.5**	**23**	**27.4**	**32**	**38.1**	**84**

Table 5 shows that 66.7% of public sector cases lie in the low-efficiency cluster of financial performance, 4.8% lie in the moderate-efficiency cluster, and 28.6% are in the high-efficiency cluster. Private sector cases are composed of 50% moderate-efficiency clusters and 47.6% high-efficiency clusters of financial performance.

To test of the impact of the insurance company's ownership type (public or private) on the fuzzy classification for financial performance efficiency, we applied the Chi-square test with the following hypotheses:

H_0: There is no relationship between the fuzzy classification for financial performance efficiency and the insurance company's ownership type.

H_1: There is a relationship.

	Value	df	Sig.
Person Chi-Square	42.834	2	0.000

The Sig. of the test is 0.000, which is less than 0.05; therefore, there is a relationship between the fuzzy classification for financial performance efficiency and the insurance company's ownership type.

Table 6 cross-tabulates the insurance companies with the fuzzy classification for efficiency of financial performance.

Table 6—Frequency and Percentage of Cases Based on Fuzzy Classification For Efficiency of Financial Performance and the Insurance Companies

Insurance Company	Efficiency of Financial Performance						Total
	Low		Moderate		High		
	No.	%	No.	%	No.	%	
Misr	13	92.9	1	7.1			14
Al-Chark	6	42.9	1	7.1	7	50	14
National	9	64.3			5	35.7	14
Suez Canal			14	100			14
El-Mohandes			4	28.6	10	71.4	14
Delta	1	7.1	3	21.4	10	71.4	14
Total	**29**	**34.5**	**23**	**27.4**	**32**	**38.1**	**84**

Table 6 shows that the best companies with the highest financial performance efficiencies are El-Mohandes and Delta. Both companies have 71.4% of its cases in high-efficiency clusters of financial performance. Misr has the worst showing with 92.9% of its cases in the low-efficiency cluster.

To test insurance companies and the fuzzy classification for the efficiency of financial performance, we applied the Chi-square test with the following hypotheses:

H_0: There is no relationship between the fuzzy classification for the efficiency of financial performance and insurance companies.

H_1: There is a relationship.

	Value	df	Sig.
Person Chi-Square	84.928	10	0.000

The Sig. of Chi-square test is 0.000, which is less than 0.05; therefore, there is a relationship between the fuzzy classification for efficiency of financial performance and insurance companies.

6.4 Multiple Discriminant Analysis

When the dependent variable has more than two values, there will be more than one regression equation. In fact, the number of regression equations is equal to one less than the number of values.

Our dependent variable is categorical and we have three fuzzy clusters; therefore, there are two dependent variables LDF_1 and LDF_2.

The independent variables represent the six factors (f_1, f_2, f_3, f_4, f_5, f_6) resulting from factor analysis for the efficiency of financial performance ratios.

Therefore, if we have three fuzzy clusters for efficiency of financial performance of the Egyptian insurance companies, then the two discriminant functions are:

$$LDF_1 = 0.580\ f_1 + 1.007\ f_2 + 0.946\ f_3 - 0.460\ f_4 + 0.195\ f_5 - 0.466\ f_6$$

$$LDF_2 = 0.675\ f_1 + 0.545\ f_2 - 0.549\ f_3 + 1.017\ f_4 + 0.412\ f_5 - 0.195\ f_6$$

6.4.1 A Goodness of Fit for Discriminant Analysis Model

There are many measures to determining the efficiency of the discriminant analysis model.

6.4.1.1 Wilks' Lambda

This measure represents the percentage of the sum of squares within the group and the total sum of squares, which are valued between zero and one. If the percentage is close to one, there is no difference between groups; alternatively, if the percentage is close to zero, there is a difference between the groups.

The value of Wilks' Lambda statistic for the two discriminant functions are 0.067 and 0.278 respectively, which indicates that there are differences among the groups of the fuzzy clusters for efficiency of financial performance.

6.4.1.2 Chi-Square Test

This measure is used to test the significance of the discriminant analysis model as follows:

Discriminant Analysis Model	**Chi-Square Value**	**df**	**Sig.**
LDF_1	212.554	12	0.000
LDF_2	100.380	5	0.000

The Sig. of the Chi-square test is 0.000, which is less than 0.05, thus indicating the significance of the discriminant analysis model and its dependability in predicting the classifications of financial performance efficiencies for the Egyptian insurance companies.

6.4.1.3 Canonical Correlation

This measure refers to the correlation between the value of the discriminant function and the independent variables in the function. For the first discriminant function, LDF_1, the canonical correlation coefficient value is 0.872; the value for the second discriminant function, LDF_2, is 0.849. Thus, there is a strong relationship between the discriminant functions and the independent variables.

6.4.1.4 Percentage of Variance

Using this measure, the first discriminant function yields 55% of variance and the second discriminant function yields 45%. Thus, the independent variables in the two discriminant functions can explain the variability in discriminant scores by 100%.

6.4.1.5 Percentage of Correct Classification

This measure applies the proposed discriminant model on our data and classifies the cases into three categories of low-, moderate-, or high-efficiency of financial performance. Table 7 yields the following results:

Table 7—Classification Results of Discriminant Model for Efficiency of Financial Performance

Original Group Membership	Predicted Group Membership						Total	
	Low		*Moderate*		*High*			
	No.	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**
Low	28	96.6	0	0	1	3.4	29	100
Moderate	1	4.3	22	95.7	0	0	23	100
High	0	0	1	3.1	31	96.9	32	100

From Table 7, we can conclude that the discriminant model for predicting the efficiency of financial performance for Egyptian insurance companies is correctly classified by 96.4%.

6.5 The Logistic Regression Model

Using the logistic regression model as a tool for predicting the financial performance efficiency of property and liability Egyptian insurance companies, yields the following results:

The estimated logistic regression models are:

$$\log it(y_2) = 0.3043 + 2.9716\ f_1 + 2.198\ f_2 - 0.9773\ f_3 + 4.2337\ f_4 + 1.323\ f_5 - 1.476\ f_6$$

$$\log it(y_3) = 0.2124 + 1.508\ f_1 - 1.5203\ f_2 - 6.311\ f_3 + 6.798\ f_4 + 0.851\ f_5 - 0.894\ f_6$$

The two equations above can be used to predict the probability that an individual company belongs to each group of financial performance efficiency (low, moderate, high). Whereas the probability that an insurance company belongs to moderate-efficiency cluster of financial performance can be written as

$$p(y=2 \mid X) = \frac{1}{(1+\exp(\log\ it(y_2)))}$$

and the probability of an insurance company belongs to high-efficiency cluster of financial performance can be written as

$$p(y=3 \mid X) = \frac{1}{(1+\exp(\log\ it(y_3)))}.$$

If the case does not belong to the moderate- or high-efficiency clusters of financial performance, then it belongs to the low-efficiency cluster.

6.5.1 A Goodness of Fit for Logistic Regression Model

6.5.1.1 Likelihood Ratio Test

This is the test of choice in logistic regression. The likelihood ratio test statistic is -2 times the difference between the log likelihoods of two models. The -2 log likelihood ratio test is approximately the chi- square distribution. Therefore, we have the following:

	Value	**Sig.**
Person Chi-Square	15.180	0.000

The Sig. of the Chi-square test is 0.000, which less than 0.05, which in turn indicates the significance of logistic regression as predictive model for determining the financial performance efficiency of the Egyptian insurance companies.

6.5.1.2 *R*-Square

The *R*-square of logistic regression model is 0.9567, which means that the independent variables $(f_1, f_2, f_3, f_4, f_5, f_6)$ in the logistic regression model can interpret the changes in the efficiency of financial performance of the Egyptian insurance companies by 95.67%.

6.5.1.3 Percent of Correctly Classification

This measure applies the proposed logistic regression model to our data and classifies the cases into three categories (low, moderate, high) of financial performance efficiencies. Table 8 shows our results.

Table 8—Classification Results of Logistic Regression Model for Efficiency of Financial Performance

Original Group Membership	Predicted Group Membership						Total	
	Low		*Moderate*		*High*			
	No.	%	No.	%	No.	%	No.	%
Low	29	100	0	0	0	0	29	100
Moderate	0	0	23	100	0	0	23	100
High	0	0	0	0	32	100	32	100

From Table 8, we can conclude that the logistic regression model predicts the of financial performance efficiencies for Egyptian insurance companies correctly 100% of the time.

Conclusions

Comparing the public and private sectors, the mean of efficiency of financial performance ratios does not vary significantly for the following ratios: return on investments, net profit to total assets, net profit to surplus, total liabilities to total assets, and underwriting expenses paid to premiums written.

Predictive variables, which represent 25 ratios measuring efficiency and financial performance data, can be reduced into six factors by using factor analysis.

Fuzzy cluster procedure indicates that the data involving the efficiency of financial performance for the Egyptian insurance companies can be classified into three groups: low, moderate, and high.

The number of financial performance cases exhibiting high efficiency is 32, or 38.1% of the total cases; the number of moderate-efficiency cases is 23, or 27.4 of total cases; and the number of low-efficiency cases represent 34.5% of total cases.

Public sector cases represent 66.7% of the low-efficiency clusters of financial performance, while private sector cases comprise 47.6% of high-efficiency clusters for financial performance. Thus, there is a relationship between the fuzzy classification of the insurance company's financial performance efficiency and its ownership type.

The best companies in efficiency of financial performance are El-Mohandes and Delta, while the worst company is Misr. There is relationship between the fuzzy classification for efficiency of financial performance and insurance companies.

Wilks' Lambda and Chi-square test indicate that discriminant analysis model is significant in determining the efficiency of financial performance of the Egyptian insurance companies.

The canonical correlation coefficients for the two discriminant functions are 0.872 and 0.849, which reflect the strong relationship between the independent variables in the model and the efficiency of financial performance.

The discriminant analysis model can predict a correct classification for the efficiency of financial performance of Egyptian insurance companies by 96.4%.

Likelihood ratio test indicates that the logistic regression model is significant for determining the efficiency of financial performance of Egyptian insurance companies.

The independent variables in the logistic regression model can interpret the efficiency of financial performance of Egyptian insurance companies by 95.67%.

The logistic regression model can predict a correct classification for the efficiency of financial performance of Egyptian insurance companies by 100%.

The results indicate that the logistic regression model can more accurately forecast the financial performance efficiencies of Egyptian insurance companies than the discriminant analysis model.

Acknowledgment

The author would like to express his appreciation to Casualty Actuarial Society and the reviewers of his submission.

References

[1] Alexander, F. E. 1999. "Cluster and clustering of childhood cancer: A review." *European Journal of Epidemiology* 15:847-852.

[2] Ambrose, Jan M. and Anne M. Carroll. 1994. "Using Best's ratings in life insurer insolvency prediction." *Journal of Risk and Insurance*. 61, no. 2:317-327.

[3] Ambrose, Jan M. and Seward J. Allen. 1988. "Best's ratings, financial ratios and prior probabilities in insolvency prediction." *Journal of Risk and Insurance* 55, no. 2:229-244.

[4] BarNiv, Ran and Hershbarger Robert A. 1990. "Classifying financial distress in the life insurance industry." *Journal of Risk and Insurance* 57, no. 1:110-136.

[5] BarNiv, Ran and McDonald James B. 1992. "Identifying financial distress in the insurance industry: A synthesis of methodological and empirical issues." *Journal of Risk and Insurance* 59, no. 4:543-573.

[6] BarNiv, Ran, John Hathorn, Abraham Mehrez, and Douglas Kline. 1999. "Confidence intervals for the probability of insolvency in the insurance industry." *Journal of Risk and Insurance* 66, no. 1:125-137.

[7] Beaver, William H. 1966. "Financial ratios as predictors of failure." *Journal of Accounting Research* 4:71-111.

[8] Beirlant, J., V. Derveaux, M. De Meyer, J. Goovaerts, E. Labie and B. Maenhoudt 1992, "Statistical risk evaluation applied to (Belgian) car insurance." *Insurance: Mathematics and Economics* 10, no. 4:289-302.

[9] Carreno, L. and Y. Jani. 1993. "A fuzzy expert system approach to insurance risk assessment using Fuzzy Clips." In *WESCON/1993 Conference Record:* 536-541

[10] Chidambaran, N. K., Thomas A. Pugel, and Anthony Saunders. 1997. "An investigation of the performance of the U.S. property-liability insurance industry." *Journal of Risk and Insurance* 64, no. 2:371-382.

[11] Cooper, Peter J. and Yvonne Y. Zheng. 2007. "Estimating the contributions of speeding and impaired driving to insurance claim cost." *Journal of Safety Research* 38, no. 1:17-23.

[12] Cormack, R. M. 1971. "A review of classification." *Journal of the Royal Statistical Society*. Series A (General) 134, no. 3:321-367.

[13] Cummins, J. David and Richard A. Derrig. 1993. "Fuzzy trends in property-liability insurance claim costs." *Journal of Risk and Insurance* 60, no. 3:429-465.

[14] Devaney, Sharon A. 1994. "The usefulness of financial ratios as predictors of household insolvency: Two perspectives." *Financial Counseling and Planning* 5:5-24.

[15] Ebanks, B., W. Kanvowski, and K. Ostaszewski. 1992. "Application of measures of fuzziness to risk classification in insurance." In Proceedings of the Fourth International Conference on Computing and Information, 290-291. Los Alamitos, California: IEEE Computer Society Press.

[16] Esteban, F. and G. Jose. 2001. "Robust logistic regression for insurance risk classification." Business

Economics Working Papers wb016413, Universidad Carlos III, Departamento de Economía de la Empresa.
[17] Fiegenbaum, A. and H. Thomas. 1990. "Strategic groups and performance: The U.S. insurance industry, 1970-84." Strategic Management Journal 11, no. 3:197-215.
[18] Harrington, Scott E. and Jack M. Nelson. 1986. "Regression-based methodology for solvency surveillance in the property-liability insurance industry." *Journal of Risk and Insurance* 53, no. 4:583-605.
[19] Hosmer, D. W. and S. Lemeshow. *Applied Logistic Regression*. 2000. 2nd ed. New York: Wiley-Interscience/Wiley Series in Probability and Statistics.
[20] Huberty, C. J. 1994. *Applied Discriminant Analysis*. New York: Wiley-Interscience/Wiley Series in Probability and Statistics.
[21] Huberty, C. J. and S. Olejnik. 2006. *Applied MANOVA and Discriminant Analysis*. New York: Wiley-Interscience/Wiley Series in Probability and Statistics.
[22] Jaccard, J. J. 2001. *Interaction Effects in Logistic Regression*. Series: Quantitative Applications in the Social Sciences. Thousand Oaks, California: Sage Publications, Inc.
[23] Jensen, R. 1971. "A cluster analysis study of financial performance of selected business firms." *The Accounting Review* 46, no. 1:36-56.
[24] Kaufman, L. and P. Rousseeuw. 1990. "Finding groups in data: an introduction to cluster analysis." New York: Wiley-Interscience/Wiley Series in Probability and Statistics.
[25] Kleinbaum, D. G., M. Klein, and E. Rihl Pryor. 2005. *Logistic Regression*. 2nd ed. New York: Springer.
[26] Lai, Gene C. and Piman Limpaphayom. 2003. "Organizational Structure and Performance: Evidence from the Nonlife Insurance Industry in Japan." *Journal of Risk and Insurance* 70, no. 4:735-757.
[27] Lazzerini, B., D. Dumitrescu, and L. C. Jain. 2000. *Fuzzy Sets* & *their Application to Clustering* & *Training*. Boca Raton, Florida: CRC Press International Series on Computational Intelligence.
[28] Lee, Suk Hun and Jorge L. Urrutia. 1996. "Analysis and prediction of insolvency in the property-liability insurance industry: A comparison of logit and hazard models." *Journal of Risk and Insurance* 63, no. 1:121-130.
[29] Lemaire, Jean. 1990. "Fuzzy insurance." *ASTIN Bulletin* 20, no.1:33-56.
[30] McLachlan, G. J. 2004. *Discriminant Analysis and Statistical Pattern Recognition*. New York: Wiley-Interscience/Wiley Series in Probability and Statistics.
[31] Menard, S. W. 2001. "Applied Logistic Regression Analysis (Quantitative Applications in the Social Sciences). 2nd ed. Thousand Oaks, California: Sage Publications, Inc.
[32] Mosley, Roosevelt. 2005. *The Use of Predictive Modeling in the Insurance Industry*. Bloomington, Indiana: Pinnacle Actuarial Resources, Inc.
[33] NCSS Statistical & Power Analysis Software. *NCSS Manual*. 2007. http://www.ncss.com.
[34] Oliveira, J. V. and W. Pedrycz, eds. 2007. *Advances in Fuzzy Clustering and its Applications*. New York: John Wiley and Sons.
[35] Peck, R., Lloyd Fisher, and John Van Ness. 1989. "Approximate confidence intervals for the number of clusters." *Journal of the American Statistical Association* 84, no. 405:184-191.
[36] Pedrycz, W. 2005. *Knowledge-Based Clustering: From Data to Information Granules*. New York: John Wiley and Sons/Wiley-Interscience.
[37] Pottier, Steven W. 1998. "Life insurer financial distress, Best's ratings and financial ratios." *Journal of Risk and Insurance* 65, no. 2:275-288.
[38] Samson, Danny. 1986. "Designing an automobile insurance classification system." *European Journal of Operational Research* 27, no. 2:235-241.
[39] Sánchez, J. and A. Gomez. 2003. "Applications of fuzzy regression in actuarial analysis." *Journal of Risk and Insurance* 70, no. 4:665-699.
[40] Sánchez, Jorge de Andrés. 2006. "Calculating insurance claim reserves with fuzzy regression." *Fuzzy Sets and Systems* 157, no. 23:3091-3108.
[41] Sánchez, Jorge de Andrés. 2007. "Claim reserving with fuzzy regression and Taylor's geometric separation method." *Insurance: Mathematics and Economics* 40, no. 1:145-163.
[42] Sato, M., Y. Sato, and L. C. Jain. 1997. *Fuzzy Clustering Models and Applications*. Heidelberg: Physica-Verlag.
[43] Shapiro, Arnold F. 2004. "Fuzzy logic in insurance." *Insurance: Mathematics and Economics* 35, no. 2:399-424.
[44] Steven, W. and W. David. 2000. "Capital ratios and property insurer insolvencies." *Journal of Risk and Insurance* 70.
[45] Tibshirani, R., Walther G. Guenther and T. Hastie. 2001. "Estimating the number of clusters in a data set via the gap statistic." *Journal of the Royal Statistical Society,* Series B (Statistical Methodology) 63, no. 2:411-423.
[46] Trieschmann, James S. and George E. Pinches. 1973. "A multivariate model for predicting financially distressed P-L insurers." *Journal of Risk and Insurance* 40, no. 3:327-338.
[47] Trieschmann, James S. and George E. Pinches. 1974. "The efficiency of alternative models for solvency

surveillance in the insurance industry." *Journal of Risk and Insurance* 41, no. 4:563-577.

[48] Van Gestel, Tony, Martens David, Baesens Bart, Feremans Daniel, Huysmans Johan and Vanthienen Jan. 2007. "Forecasting and analyzing insurance companies' ratings." *International Journal of Forecasting* 23, no. 3:513-529.

[49] Verrall, R. and Y.H. Yakoubov. 1998. "A fuzzy approach to grouping by policyholder age in general insurance." Actuarial Research Paper no. 104, Department of Actuarial Science and Statistics, City University, London.

[50] Wagstaff, A. and M. Lindelow. 2008. "Can insurance increase financial risk? The curious case of health insurance in China." *Journal of Health Economics* (forthcoming) (accepted manuscript, available online 9 February 2008).

[51] Wikipedia, the Free Encyclopedia. "Cluster Analysis." http://en.wikipedia.org/wiki/Data_clustering (accessed 2007-2008).

[52] Yeo, A., K. Smith, J. Robert, R. Willis and M. Brooks. 2001. "Clustering technique for risk classification and prediction of claim costs in the automobile insurance industry." *International Journal of Intelligent Systems in Accounting, Finance* & *Management* 10, no. 1:39-50.

[53] Young, Virginia R. 1996. "Insurance rate changing: A fuzzy logic approach." *Journal of Risk and Insurance* 63, no. 3:461-484.

[54] Young, Virginia R. 1997. "Adjusting indicated insurance rates: fuzzy rules that consider both experience and auxiliary data." *Proceedings of the Casualty Actuarial Society* 84:734-765.

[55] Zheng, Yvonne Y., Peter J. Cooper, and C.B. Dean. 2007. "Modeling the contribution of speeding and impaired driving to insurance claim counts and costs when contributing factors are unknown." *Journal of Safety Research* 38, no. 1,:25-33.

Biography of the Author

Osama Hanafy Mahmoud is an assistant professor in the department of mathematics, statistics, and insurance at the Sadat Academy for Management Sciences in Egypt. He is also an assistant professor in the department of statistics and quantitative methods in the College of Management Sciences & Planning at King Faisal University in Saudi Arabia. Professor Mahmoud can be reached at K.F.U, PO Box 1760, Al Hofuf, Al HASSA 31982, Saudi Arabia; (E- Mail) Oshanafy@hotmail.com.