View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by CLoK

University of Central Lancashire

Article

On the applicability of credit scoring models in Egyptian banks

Abdou, Hussein, El-Masry, Ahmed and Pointon, John

Available at http://clok.uclan.ac.uk/25009/

Abdou, Hussein ORCID: 0000-0001-5580-1276, El-Masry, Ahmed and Pointon, John (2007) On the applicability of credit scoring models in Egyptian banks. Banks and Bank Systems, 2 (1).

It is advisable to refer to the publisher's version if you intend to cite from the work.

For more information about UCLan's research in this area go to http://www.uclan.ac.uk/researchgroups/ and search for <name of research Group>.

For information about Research generally at UCLan please go to http://www.uclan.ac.uk/research/

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the http://clok.uclan.ac.uk/policies/

CLoK
Central Lancashire online Knowledge
www.clok.uclan.ac.uk

CLoK
University of Central Lancashire's Institutional Repository

"On the Applicability of Credit Scoring Models in Egyptian Banks"

AUTHORS	Hussein Abdou Ahmed El-Masry John Pointon
ARTICLE INFO	Hussein Abdou, Ahmed El-Masry and John Pointon (2007). On the Applicability of Credit Scoring Models in Egyptian Banks. *Banks and Bank Systems*, *2*(1)
JOURNAL	"Banks and Bank Systems"
FOUNDER	LLC "Consulting Publishing Company "Business Perspectives"

NUMBER OF REFERENCES	NUMBER OF FIGURES	NUMBER OF TABLES
0	**0**	**0**

© The author(s) 2018. This publication is an open access article.

BUSINESS PERSPECTIVES

businessperspectives.org

ON THE APPLICABILITY OF CREDIT SCORING MODELS IN EGYPTIAN BANKS

Hussein Abdou[*], Ahmed El-Masry[**], John Pointon[***]

Abstract

Credit scoring is regarded as a core competence of commercial banks during the last few decades. A number of credit scoring models have been developed to evaluate credit risk of new loan applicants and existing loan clients. The main purpose of the present paper is to evaluate credit risk in Egyptian banks using credit scoring models. Three statistical techniques are used: discriminant analysis, probit analysis and logistic regression. The credit scoring task is performed on one bank's personal loans data-set. The results so far revealed that all proposed models gave a better average correct classification rate than the one currently used. Also both type I and type II errors had been calculated in order to evaluate the misclassification costs.

Key words: Credit scoring models; Discriminant analysis; Probit analysis; Logistic regression; Classification; Egyptian banks.
JEL Classification: G21, G32.

1. Introduction

Latterly, credit risks have become one of the most important financial topics of interest, especially in the banking sector. The role of credit risks has changed dramatically over the last ten decades, from passive automation to a strategic device.

The process of credit risk evaluation has the interest of many researchers nowadays. Recently, bankers have come to realise that banking operations affect and are affected by the natural environment and that consequently the banks might have an important role to play in helping to raise environmental standards. Although the environment presents significant risks to banks, in particular environmental credit risk, it also perhaps presents profitable opportunities (Thompson, 1998).

Making a decision about accepting or rejecting a client's credit can be supported by judgemental techniques and/or credit scoring models. The judgemental techniques rely on the knowledge and both past and present experience of credit analysts, who evaluate the required requisites, such as the personal reputation of a client, the ability to repay credit, guarantees and client's character. Due to the rapid increase in fund-size invested through credit granted by Egyptian banks, and the need for quantifying credit risk, financial institutions including banks have started to apply credit scoring models.

The structure of the banking system varies from country to country. In the Egyptian environment the structure includes[1]: First, public sector banks (7 banks). Second, private and joint venture banks (28 banks). Third, branches of foreign banks (7 banks). Forth, branches ceased its operations (9 banks)[2] (See for more details: http://www.cbe.org.eg/links.htm).

[*] Plymouth Business School, UK.
[**] Plymouth Business School, UK.
[***] Plymouth Business School, UK.

[1] Before 16 October 2006 the Egyptian banking structure consisted of: commercial banks (28 banks), comprising public sector banks (4 banks) and private & joint venture banks (24 banks); and secondly, business & investment banks (31 banks), comprising private & joint venture banks (11 banks) and branches of foreign banks – off-shore banks (20 banks). In addition, there are also specialised banks (3 banks), namely the Egyptian Industrial Development Bank, the Arab Egyptian Real Estate Bank and Principal Bank for Development and Agriculture Credit. Egyptian banks abroad are not included, also two banks established under private laws and are not registered with Central Bank of Egypt; namely, Arab International Bank, and Nasser Social Bank (Central Bank of Egypt, 2003/2004).
[2] The board of the CBE agreed to cancel two banks, Jammal Trust Bank and Rafidain Bank, from its record.

© Hussein Abdou, Ahmed El-Masry, John Pointon, 2007.

Since most banks in Egypt are currently using judgemental techniques, it is important to review judgemental techniques versus credit scoring models. Sullivan (1981) argues that, in a judgemental risk evaluation process, each credit application and the information contained therein are evaluated individually by an employee of the creditor. The success of a judgemental system depends on the experience and common sense of the credit evaluator. Otherwise, in a credit scoring model, some creditors have used their historical experience with debtors to derive a quantitative model for the segregation of acceptable and unacceptable credit applications. With a credit scoring system, a credit application is processed mechanically and all credit decisions are made consistently. The scoring system is based on the addition or subtraction of a statistically derived number of points relating to the applicant's credit score on the basis of responses given to a set of predictor variables, such as time on a job or the number of credit sources used. Given a statistically derived cut-off credit score, a creditor can thus separate the acceptable from the unacceptable credit applicants. On the other hand, credit scoring has been criticized because statistical problems with the data used to derive the model frequently violate the assumptions of the statistical technique used to derive the points (Sullivan, 1981: 9.17).

It is also pointed out that some of the variables used in a credit scoring system may have the effect of social discrimination, although to the statistician the variable may appear to be neutral. Finally, the credit scoring model is derived by analysis of the characteristics of the customers who were once granted credit by the creditor for whom the system is derived. The characteristics of the part of the population to which the credit grantor has not granted credit are not directly considered. Thus the scoring system may provide biased results when it is applied to new credit applicants. Despite the criticism of credit scoring models, these models can be regarded as one of the most successful models used in the field of business and finance (Sullivan, 1981: 9.17).

Credit scoring is the set of decision models and their underlying techniques that aid lenders in the granting of consumer credit. These techniques assess, and therefore help to decide, who will get credit, how much credit they should get, and what operational strategies will enhance the profitability of the borrowers to the lenders (Long, 1973; and Thomas et al., 2002).

Discriminant analysis and linear regression are widely-used statistical techniques, as evidenced in the literature that follows. The other methods are: logistic regression, probit analysis, mathematical programming, non-parametric smoothing methods, Markov chain models, expert systems, neural networks, genetic algorithms and others (Hand and Henley, 1997). For such a new banking environment, it would see appropriate, as a first step, to investigate some of the conventional techniques such as discriminant analysis, probit analysis and logistic regression.

Indeed, discriminant analysis and logistic regression are still used in building and developing credit scoring models (Sarlija et al., 2004; Hand and Henley, 1997; and Caouette et al., 1998). Generally, the best technique for all data sets does not exist. Therefore, the main thrust of this paper is to explore credit scoring models to evaluate credit risk in the banking sector in Egypt, in terms of a case study, using some statistical techniques such as discriminant analysis, probit analysis and logistic regression. Discussion with banking officials would suggest that most banks in Egypt are using judgemental techniques in their evaluation process. We are examining integrated models for the evaluation of consumer credit risks in the banking sector in Egypt; especially since credit scoring models have undergone a noticeable success in different environments in Europe and the US, taking into account all requirements for the proposed models according to the nature of the Egyptian environment.

The empirical results, with a 0.50 cut-off point, reveal that an 86.75% average correct classification rate was found using discriminant analysis, and with a stepwise discriminant approach, nine significant predictor variables are selected in the final model and found an 86.92% as the average correct classification rate. An 87.78% average correct classification rate for probit analysis was found. Moreover, an 87.26% average correct classification rate was observed after excluding the insignificant variables. Using logistic regression, it was found that the average correct classification rate is 88.30%, and 87.95% after excluding the insignificant variables. In general, all models gave better correct classification rates than the currently used system (74.53% of all accepted loans which did not lead to default, i.e. 433/581). Misclassification costs are also investigated in this paper; since the costs associated with type I errors differ from those associated with type II errors.

This paper is organized as follows: section two discusses the literature review. Section three details the research methodology and data collection. Section four explains the results. Finally, section five concludes the study results and suggests the area for the future researches.

2. Literature review

Credit scoring was one of the earliest financial risk management tools developed. Its use by US retailers and mail-order firms in the 1950s was coetaneous with the early applications of portfolio analysis to manage and diversify the risk inherent in investment portfolios. In addition, credit scoring could claim to be the grandfather of data mining because it was one of the earliest uses of data on consumer behaviour (Thomas et al., 2002).

The objective of credit scoring models is to assign loan customers to either good credit or bad credit (Lee et al., 2002). Therefore, scoring problems are related to classification analysis (Anderson, 2003; Hand, 1981; and Lee et al., 2002). Classification models for credit scoring are used to categorize new applicants as either accepted or rejected with respect to their characteristics, such as, marital status, age, and income (Chen and Huang, 2003). At the same time, this suits the Egyptian environment, with perhaps the addition of other variables, such as corporate guarantee, monthly salary and education.

The credit scoring model is one of the most successful applications of research modelling in finance and banking, and the number of scoring analysts in the industry is constantly increasing. Yet because credit scoring does not have the same lustre as the pricing of exotic financial derivatives or portfolio analysis, the literature on the subject is very limited. However, credit scoring has been vital in allowing the phenomenal growth in consumer credit over the last four decades. Without an accurate and automatically operated risk assessment tool, lenders of consumer credit could not have expanded their loan books in the way they have (Lewis, 1992; Bailey, 2001; Mays, 2001; Thomas et al., 2002; Bluhm et al., 2003; and Siddiqi, 2006).

Possibly the earliest use of applying multiple discriminant analysis to credit scoring is the work by Durand (1941), who examined car loan applications. A well-known application in corporate bankruptcy prediction is one by Altman (1968), who developed the first operational scoring model based on five financial ratios, taken from eight variables from corporate financial statements. He produced a Z-Score, which is a linear combination of the financial ratios.

The evaluation of new consumer loans is one of the most important applications of credit scoring models and it has attracted some attention in the last few decades (Steenackers and Goovaerts, 1989; and Sarlija et al., 2004). Some studies focus on existing consumer loans rather than new loan applications (Orgler, 1971; and Kim and Sohn, 2004).

Statistical techniques, such as discriminant analysis, regression analysis, probit analysis and logistic regression, used in building the scoring models have been examined (Orgler, 1971; Boyes, et al., 1989; Steenackers and Goovaerts, 1989; Greene, 1998; Banasik et al., 2001; and Sarlija et al., 2004). There have also been case studies of building credit scoring models (Leonard, 1995; Banasik et al., 2001; Lee et al., 2002; and Lee and Chen, 2005).

On the one hand, the use of only two groups of customer credit, either "good" or "bad" as it has been used here, is appropriate within such a new environment, such as the Egyptian banking sector, to credit scoring models, and is still one of the most important assortments in credit scoring applications (Orgler, 1971; Boyes et al., 1989; Banasik et al., 2001; Lee et al., 2002; and Kim and Sohn, 2004). On the other hand, the use of three groups of consumer credit became one of the approaches for classification in credit scoring models. Some have used "good" or "bad" or "refused" (Steenackers and Goovaerts, 1989), whilst others have used "good" or "poor" or "bad" (Sarlija et al., 2004). Otherwise, the probit analysis (Guillen and Artis, 1992; Banasik et al., 2003; and Greene, 1998) has been used in building credit scoring models beside other statistical techniques.

It is important for new users to apply the most appropriate technique(s) for the array of methods available, bearing in mind comparisons between different methods (Guillen and Artis, 1992; Desai et al., 1996; Hand and Henley, 1997; Baesens et al., 2003; Chen and Huang, 2003; and Ong et al., 2005), and the emphasis on a dichotomous variable of "good" and "bad" (Guillen and Artis, 1992; Desai et al., 1996; Hand and Henley, 1997; Banasik et al., 2003; Chen and Huang, 2003; and Yang et al., 2004), in building the scoring models, especially for the new users to credit scoring models.

Lim and Sohn (2007) argue that using existing models is quite troublesome to discriminate the creditability of borrowers with high default risks in the middle of the repayment term. However, with the cluster-based dynamic scoring models, the lender can identify the individual credibility at earlier stage of loan period without loosing its accuracy.

In general, there is no overall best statistical technique/method used in building credit scoring models, for what is best depends on the details of the problem, the data structure, the characteristics used, the extent to which it is possible to segregate the classes by using those characteristics, and the objective of the classification (Hand and Henley, 1997). Most studies that made a comparison between different techniques found that, first, most recent/advanced statistical techniques such as neural networks and fuzzy algorithms are better than the traditional ones; second, there is no apparent difference between different statistical techniques in terms of the percentage of average correct classification or hit rate. This sometimes depends on the original group that is used to compute the correct classification, depending on "bad" or "good and bad" together (Desai et al., 1996; Blochlinger and Leippold, 2006; Hoffmann et al., 2007). However, the more simple classification techniques, such as linear discriminant analysis and logistic regression, also have a very good performance in this context, which is in majority of the cases not statistically different from other techniques (Baesens et al., 2003).

The chosen environment will be the Egyptian banking sector, in which no study (in the best of our knowledge) has investigated the use of sophisticated statistical appraisal techniques in credit scoring. Indeed, from the review of literature to date, no studies were found in Egypt in covering credit scoring techniques. Therefore, we intend to cover this gap, which was found in the Egyptian banking sector.

3. Methodology and Data Collection

In this part, three models used in building credit scoring are described first. The first model is the discriminant analysis model (DA), which was first proposed by Fisher (1936) as a discrimination and classification technique. The second model is the probit analysis model (PA), which is also usually used with other statistical techniques for the purpose of comparing the results. Newly, the logistic regression model (LR), unlike other statistical techniques, can suit different kinds of distribution functions and is more suitable for credit scoring problems. Later, the data collection method and the identification of variables will be discussed.

3.1. Credit Scoring Models

3.1.1. Discriminant Analysis

However, as to the statistical assumptions implicit in implementation, DA requires the data to be independent and normally distributed. Consequently, the general formula of DA is as follows:

$Z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots\ldots\ldots\ldots\ldots + \beta_n X_n,$

where Z represents the discriminant (zed) score, α is the intercept term, and β_i represents the respective coefficient in the linear combination of explanatory variables, X_i, for i = 1 to n (Lee et al., 2002).

Specifically, the DA model assumes that (Desai et al., 1996):

- the predictor variables are measured on an interval scale;
- the covariance matrices of the predictor variables are equal for the two groups; and
- the predictor variables follow a multivariate normal distribution.

3.1.2. Probit Analysis

PA is a technique that finds coefficient values, such that this is a probability of a unit value of a binary coefficient. As such Probit means "probability unit". Under a probit model, a linear combination of the independent variables is transformed into its cumulative probability value from a normal distribution. The method requires finding value for the coefficients in this linear combination, such that this cumulative probability equals the actual probability that the binary outcome is one, thus:

$Prob\ (y = 1) = \Phi\ (\alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots\ldots\ldots\ldots\ldots + \beta_n X_n)$,

where y is the zero-one binary outcome for a given set of value. Φ is the value from the cumulative normal distribution function. A is the intercept term, and β_i represents the respective coefficient in the linear combination of explanatory variables, X_i , for $i = 1$ to n.

PA is used as an alternative to LR. Early in the 1930s the term "Probit" has been developed which stands for probability unit (Maddala, 2001; and Pindyck and Rubinfeld, 1997).

3.1.3. Logistic Regression

LR is a widely used statistical modelling technique in which the probability of a dichotomous outcome (zero or one) is related to a set of potential predictor variables in the form:

$\log[p/(1\text{-}p)] = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots\ldots\ldots\ldots\ldots + \beta_n X_n$,

where p is the probability of the outcome of interest, α is the intercept term, and β_i represents the respective coefficient in the linear combination of explanatory variables, X_i , for i = 1 to n. The dependent variable is the logarithm of the odds, {Log [p/ (1-p)]}, which is the logarithm of the ratio of two probabilities of the outcome of interest (Lee et al., 2002).

Given the set of explanatory variables, the probability of a value of one for the dichotomous outcome is (Desai et al., 1996):

$$Z = \frac{1}{1 + e^{-Z}} ,$$

where

Z = the probability that the dichotomous outcome is one; and

$Z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots\ldots\ldots\ldots\ldots + \beta_n X_n$.

Thus, the objective of a logistic regression model in credit scoring is to determine the conditional probability of a specific observation belonging to a class, given the values of the independent variables of that credit applicant (Lee and Chen, 2005).

PA tends to be used as alternative to LR, although LR is more suited to dichotomous testing. Comparing LR with DA, the LR does not necessarily require the assumptions of DA. One advantage of DA is that the ordinary least square estimation procedure can be implemented to estimate the coefficient of the linear discriminant function, whereas the maximum likelihood method is required for the estimation of logistic regression models. Another advantage of DA over logistic regression is that prior probabilities and misclassification costs can easily be incorporated into the DA approach[1] (Desai et al., 1996). Moreover, both DA and LR have been widely used in business, finance, science, and customer behaviour (Lee et al., 2002).

---

[1] For these reasons we intend to use both techniques in this paper.

3.2. Data Collection and Proposed Variables

In order to build the proposed three credit scoring models, a personal loans dataset was provided by one of the biggest commercial banks in Egypt. This consists of 581 personal loans with 433 good loans and 148 bad loans. It should be emphasized that this dataset is pertinent because of the large number of bad loans (25.47%) with good loans (74.53%). Each bank customer in this dataset is linked to 20 independent variables (see Table 1 for details), in addition to the dependent variable, which is loan quality explained by two values, good/paid = 1 and bad/defaulted = 0. Some variables had identical values for all cases and hence were excluded, e.g. loan duration was four years in all cases, and all customers had a credit card.

Table 1

List of variables used in building the proposed credit scoring models

Variable/Description	Code	Unit	Comment
Loan Amount	LOAN AMO	No.	-
Loan Duration	-	-	Loan duration is 4 years in all cases in this sample.
Company	COMP	10, 01, 00	10 = Public sector, 01 = Local private sector, 00 = Multinational company.
Branch	-	-	The bank has a branch to serve and collect instalments (i.e. clients work or live in a very remote area that there is no branch in the city).
Sex	SEX	0, 1	0 = Male, 1 = Female
Marital Status	MAR STA	0, 1	0 = Married, 1 = Single
Age	AGE	Years	Clients ages from 25 to 59 years.
Salary/Monthly Income	SALA	No.	-
Additional Income	ADD INC	0, 1	0 = N/A, 1 = Suitable
House Owned or Rented	HOR	0, 1	0 = Rented, 1 = Owned
House Rent > Loan Tenure or House Rental Period	-	-	The client must have a rent contract for 4 years or higher to be greater than loan tenure (4 years).
Home Telephone	TELE	0, 1	0 = N/A, 1 = Ok confirmed (land line).
Utility Bill	-	-	Clients must have a utility bill not less than 6 months.
Title/Position	-	-	It means the occupation of customers: workers are less grade than white collar, workers are not accepted.
Education Level	EDU	0, 1	0 = University, 1 = Higher education 100% university or higher, it is a must.
Loans From Other Banks	LFOB	0, 1	0 = N/A, 1 = Nil
Relation With Other Banks	-	-	Through an investigation report from the central bank of Egypt (provides the client's history).
Credit Card Status	-	-	All clients have valid credit card(s).
Corporate Guarantee	COR GUAR	0, 1	0 = No, 1 = Ok from creditable company. There is no such a default with a client has a corporate guarantee.
Other Guarantors	-	-	If required.
Loan Quality	LOAN QUA	0, 1	0 = Default/Bad credit, 1 = Paid/Good credit

Selected variables for the proposed models were reduced to 12 variables, as shown in Table 2. In addition, all clients must have an investigation report from the Central Bank of Egypt, which provides a comprehensive history of the clients' dealings with all banks in Egypt.

Table 2

List of predictor variables proposed in building the credit scoring models

Variables/Description
X_1 Loan Amount*
X_2 Loan Duration
X_3 Company*
X_4 Branch
X_5 Sex*
X_6 Marital Status*
X_7 Age*
X_8 Salary/Monthly Income*
X_9 Additional Income*
X_{10} House Owned or Rented*
X_{11} House Rent > Loan Tenure
X_{12} Home Telephone*
X_{13} Utility Bill
X_{14} Title/Position
X_{15} Education Level*
X_{16} Loans From Other Banks*
X_{17} Relation With Other Banks
X_{18} Credit Card Status
X_{19} Corporate Guarantee*
X_{20} Other Guarantors

* Variables finally selected in the four credit scoring models.

4. Results

In order to run the proposed models, STATGRAPHICS Plus 5.1 and SPSS Software (SPSS 11.5) were used in this study. The detailed credit scoring results using the above-mentioned three modelling techniques can be summarized as follows. Because of the high correlation between the loan amount and monthly salary, 0.963, an Orthogonalisation test has been used to keep the effect of both in the proposed models because of their potential importance. The revised correlation, after running the test, was 0.269; all other variables had correlations within an acceptable range.

4.1 Discriminant Analysis

DA credit scoring models were designed to develop a set of discriminating functions, which can help predict the dependent variable. All the 12 predicted variables were entered. The one discriminating function with a P-value of 0.0000 was statistically significant at the 95% confidence level.

Table 3

Classification results using the DA

Observed group	Predicted group	
	1 Good	0 Bad
1 Good	372	61
0 Bad	16	132

Average correct classification rate: 86.75%.
Cut-off point 0.50.

From the results revealed in Table 3, it can be observed that the average correct classification rate is 86.75%, depending on 0.5 prior probabilities for groups. Again a stepwise discriminant approach (Neter et al., 1996; Johnson and Wichern, 2002; and Lee et al., 2002) was adopted in building the DA scoring model (which we call DA_1).

The stepwise approach was run on a forward basis, entering at each step the variable that minimizes the overall Wilks' lambda. The minimum partial F to enter was 3.84, and the minimum partial F to remove was 2.71. Prior probabilities were used treating all groups equally, and the covariance matrix was applied 'within groups'. Nine significant predictor variables are selected in the final model (discriminant function), LOAN AMO, COR GUAR, TELE, LFOB, AGE, MAR STA, EDU, HOR, and SALA. From Table 4, 86.92% was observed as the average correct classification rate.

Table 4

Classification results using the DA_1

Observed group	Predicted group	
	1 Good	0 Bad
1 Good	372	61
0 Bad	15	133

Average correct classification rate: 86.92%.
Cut-off point 0.50.

4.2. Probit Analysis

PA credit scoring models were developed to describe the relationship between the dependent variable (LOAN QUA) and 12 independent variables. Because the P-value for the model in the analysis of deviance table (Appendix A) is less than 0.01, there is a statistically significant relationship between the variables at the 99% confidence level. In addition, the P-value for the residuals is greater than or equal to 0.10, indicating that the model is not significantly worse than the best possible model for this data at the 90% or higher confidence level, as it is shown in Appendix A.

All selected variables were significant at the 95% confidence level except three variables: ADD INC, SEX, and COMP[1]. But because of their potential importance we kept them in the model. Table 5 reveals an 87.78% average correct classification rate for this model using a 50% cut-off point. Nevertheless, the highest correct classification per cent was found using a 65% cut-off point, which is 89.33%.

[1] In addition to HOR with a P-value of 0.1002, but after excluding the three variables became significant with a P-value of 0.0179.

Table 5

Classification results using the PA

Observed group	Predicted group	
	1 Good	0 Bad
1 Good	407	26
0 Bad	45	103

Average correct classification rate: 87.78%.
Cut-off point 0.50.

Hence we ran the model again, without ADD INC, SEX and COMP (calling this the PA_1 model). All included variables were significant, and an 87.26% average correct classification rate was observed with a cut-off of 50% as it is shown in Table 6. The highest average correct classification rate at 88.81%, using a 60% cut-off point, was found.

Table 6

Classification results using the PA_1

Observed group	Predicted group	
	1 Good	0 Bad
1 Good	403	30
0 Bad	44	104

Average correct classification rate: 87.26%.
Cut-off point 0.50.

4.3. Logistic Regression

Table 7 summarizes the results of the LR credit scoring model, using the original 12 predictor variables. It can be observed that the average correct classification rate was 88.30% with a 0.5 cut-off point. Because the P-value (see Appendix A) for the model is less than 0.01, there is a statistically significant relationship between the variables at the 99% confidence level. In addition, the P-value for the residuals is greater than or equal to 0.10, indicating that the model is not significantly worse than the best possible model for this data at the 90% or higher confidence level. The highest correct classification rate was 89.85% with a 0.60 cut-off point.

Table 7

Classification results using the LR

Observed group	Predicted group	
	1 Good	0 Bad
1 Good	407	26
0 Bad	42	106

Average correct classification rate: 88.30%.
Cut-off point 50%.

As shown in Figure 1, there are many cases of a high probability prediction of good credit, which were confirmed as true. Where the prediction probability exceeds 0.60 there are a few false results, i.e. bad credits; and vice versa, i.e. for probabilities of good credits less than 0.60 there are more false results, than true results, i.e. more bad credits associated with low predictions of good credits, than good credits associated with predictions of good credits.

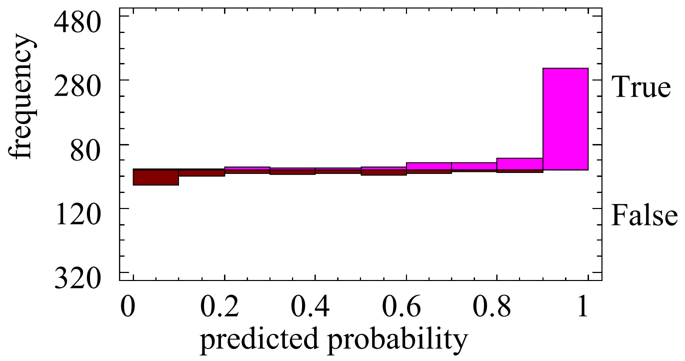

Fig. 1. Model Predictions for LOAN QUA

Prediction capability for LOAN QUA, describes the relationship between different cut-off points and the per cent correctly classified. As shown in Figure 2, the middle line refers to the true correctly classified. The highest line at the lower cut-off rates is the true correctly classified set, while the lowest line at the lower cut-off rates refers to the falsely classified set.

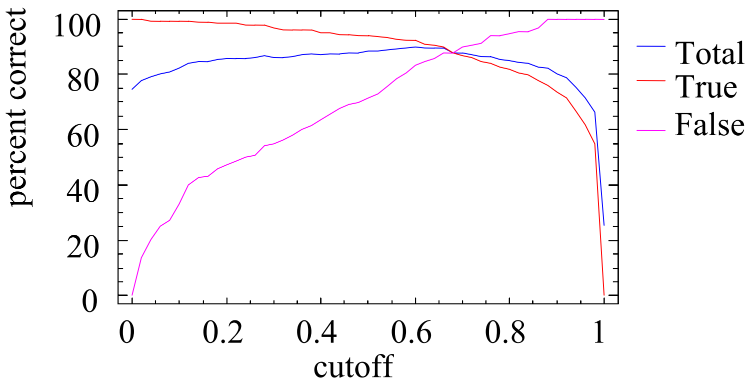

Fig. 2. Prediction Capability Plot for LOAN QUA

Actually, three variables were not significant at the 95% confidence level: ADD INC, SEX, and COMP[1]. The model was run again (which we called model LR_1) without ADD INC, SEX and COMP; all predictor variables were significant at the 95% confidence level. The average correct classification rate as it is shown in Table 8 was 87.95% with a 0.50 cut-off point, and 89.16% with a 0.60 cut-off point[2].

---

[1] In addition to HOR with a P-value of 0.1695 but after excluding just the ADD INC, the P-value of HOR became 0.0429 and 0.0275 after excluding the ADD INC, SEX and COMP.

[2] Appendix B summarizes the PA, PA1, LR, LR1 different cut-offs, and their average correct classification rates (this option was not available using discriminant analysis, the standard cut-off being 0.50 only in SPSS 14.0 and STATGRAPHICS Plus 5.1).

Table 8

Classification results using the LR_1

Observed group	Predicted group	
	1 Good	0 Bad
1 Good	406	27
0 Bad	43	105

Average correct classification rate: 87.95%.
Cut-off point 50%.

4.4. Comparison of results of different credit scoring models[1]

Since the average correct classification rate became an important criterion/tool in evaluating the classification capability of the scoring models, it was important to compare the different models' results. The classification results for all proposed models are compared in order to evaluate these models. Table 9 summarizes the average correct classification rate results for DA, DA_1, PA, PA_1, LR and LR_1. It can be concluded from Table 9 that both LR and LR1 have the highest average correct classification rates. The probability of a good loan from an accepted loan application can be derived from the LR model output.

Table 9

Comparing credit scoring results for the proposed models

	Credit scoring results		
Credit scoring model	(0-0)	(1-1)	ACC rate*
DA	89.19% (132/148)	85.91% (372/433)	86.75%
DA_1	89.86% (133/148)	85.91% (372/433)	86.92%
PA	69.59% (103/148)	94.00% (407/433)	87.78%
PA_1	70.27% (104/148)	93.07% (403/433)	87.26%
LR	71.62% (106/148)	94.00% (407/433)	88.30%
LR_1	70.95% (105/148)	93.76% (406/433)	87.95%

* the average correct classification rate calculated using 0.50 cut-off point.

Besides, LR has the highest average correct classification rate (89.85%, see Appendix B) above all proposed models with a 60% cut-off point. In general, all models gave better correct classification rates than the currently used system/model (74.53% of all accepted loans which did not lead to default, i.e. 433/581).

For the purpose of comparing results of all models developed in this study, and in order to evaluate the overall credit scoring capability and effectiveness, the misclassification costs have been taken into account, beside the average correct classification rate, in order to find the minimum expected misclassification cost in a credit scoring model (West, 2000).

The following equation is used in computing the estimated misclassification cost:

$$\text{Cost} = C\,(\text{bad/good}) \times P\,(\text{bad/good}) \times \pi_1 + C\,(\text{good/bad}) \times P\,(\text{good/bad}) \times \pi_0$$

where, C (bad/good) i.e. C (predicted bad/actually good) and C (good/bad) i.e. C (predicted good/actually bad), are the corresponding misclassification costs of both type I and type II errors.

[1] The models compared in this section depend on the observed results, using a 0.50 cut-off point only.

P (bad/good) and *P* (good/bad) measure the probabilities of type I and type II errors. π_1 and π_0, are the prior probabilities of good and bad respectively (West, 2000).

Lee and Chen (2005) state that it is complicated task to estimate the misclassification costs, as valid prediction might not be available. However, it is generally believed in a credit scoring application that the costs associated with both type I and type II errors are significantly different. Generally, the misclassification cost associated with a type II error is much higher than the misclassification cost associated with a type I error.

West (2000) noted that Dr Hofmann, who compiled his German credit data, reported that the ratio of misclassification costs associated with type II and type I is (5:1).

In this paper, this relative cost ratio will be used to calculate the estimated misclassification cost for the proposed models. The prior probabilities of good and bad credit are set as 74.53% and 25.47% respectively, using the ratio of good and bad credit in the Egyptian data-set.

Table 10 concludes the type I[1], type II[2] errors and the estimated misclassification costs for all proposed models. In general, the misclassification errors associated with type II are higher than those associated with type I, which is also true in other case studies based on credit card and housing loans datasets (Lee et al., 2002; and Lee and Chen, 2005).

Table 10

Type I, Type II errors and estimated misclassification costs for the proposed models

Credit scoring model	Error results		Estimated misclassification cost
	Type I	Type II	
DA	14.09% (61 / 433)	10.81% (16/148)	0.24268
DA_1	14.09% (61 / 433)	10.14% (15/148)	0.23415
PA	6.00% (26 / 433)	30.41% (45/148)	0.43199
PA_1	6.93% (30 / 433)	29.73% (44/148)	0.43026
LR	6.00% (26 / 433)	28.38% (42/148)	0.40614
LR_1	6.24% (27 / 433)	29.05% (43/148)	0.41646

Our results are consistent with those using probit and logistic models, namely, PA, PA_1, LR and LR_1, while the discriminant models did not agree with them. The first two models, DA and DA_1, predicted good credits for bad customers (type II) much lower than the other models did. The type I errors in the first two models are higher than the type II errors. By contrast, PA, PA_1, LR and LR_1 predicted bad credits for good customers (type I) much lower than the DA and DA_1.

Furthermore, where the type I error rate exceeds the type II error rate, as in the case of DA and DA_1, the lower misclassification cost at 0.23415 is for DA_1. Also, we know that the average correct classification rate criterion led to selecting DA_1 at 86.92% (see Table 9). Correspondingly, where the type II error rate exceeds the type I error rate, as for PA, PA_1, LR and LR_1, the lowest misclassification cost at 0.40614 is for LR. This is also the chosen model between PA, PA_1, LR and LR_1, for LR has the highest correct classification rate at 88.30% (see Table 9).

Comparing all the techniques, the lowest misclassification cost criterion leads to selecting DA_1, which is the stepwise discriminant analysis, with a minimum cost of 0.23415. However, this does not provide the highest average correct classification rate, which was 88.30% for LR.

---

[1] Good credit is misclassified as bad credit.
[2] Bad credit is misclassified as good credit.

4. Conclusion and Area of Future Research

Within a competitive environment for financial institutions, including banks, credit scoring techniques have become one of the most important tools currently used in the credit risk evaluation of loans. Besides, credit scoring is regarded as one of the basic applications of misclassification problems that have attracted more and more attention during the past decades. This study presents an evaluation of personal loans to help strengthen the credit risk evaluation process in the Egyptian banking sector using three credit scoring statistical techniques: DA, PA, and LR.

The ranking of the models varied according to the decision criterion. Using the highest average correct classification rate, LR is preferred, whereas using the lowest estimated misclassification cost, DA_1 is the best model. Further the final choice depends on the bank's decision maker's viewpoint. In other words, what are they looking for? Is it avoiding the misclassification cost and in particular the type II error? (in this case they have to select the model with the lowest cost). Or they might know from a market study that they have a strong clientele (in this case they should select the model with the highest average correct classification rate).

It needs to be observed that the analysis undertaken here has been based on loan applications that have been accepted, some of which later proved to be bad. Of course, some of the initially rejected applications may have led to recommendations of acceptance, if our scoring models had been employed. Nevertheless, the bank's own alternative systems of scoring may have filtered out many but not all applications that would also have been rejected by the statistical models. Indeed statistical credit scoring models can complement judgemental techniques. In this study we have used an in-sample data-set, i.e. the whole sample in producing the results. Subsequent applications to new loan data may reveal less accurate predictions.

Some of the predictor variables have not normally been used in published studies of credit scoring models, for example: corporate guarantee, branch, and loans from other banks. There are particularly appropriate within the Egyptian environment.

Future studies should aim to use advanced statistical scoring techniques such as neural networks, besides the traditional scoring models which were used in the current study, and perhaps integrated with other techniques such as, genetic algorithms and fuzzy discriminant analysis. In addition to this, the plan is to collect more data and employ more variables that might increase the accuracies of the scoring models. Finally, future research would use more than one bank's data-set for the purpose of generalizing the results.

References

1. Altman, E.I. 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, XXIII: 589-609.
2. Anderson, T.W. 2003. *An Introduction to Multivariate Statistical Analysis.* New York: Wiley-Interscience.
3. Baesens, B., T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen. 2003. Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring. *Journal of the Operational Research Society*, 54: 627-635.
4. Bailey, M. 2001. *Credit Scoring: The Principles and Practicalities.* Kingswood, Bristol: White Box Publishing.
5. Banasik, J., J. Crook, and L. Thomas. 2001. Scoring by Usage. *Journal of the Operational Research Society*, 52: 997-1006.
6. Banasik, J., J. Crook, and L. Thomas. 2003. Sample Selection Bias in Credit Scoring Models. *Journal of the Operational Research Society*, 54: 822-832.
7. Bluhm, C., L. Overbeck, and C. Wagner. 2003. *An Introduction to Credit Risk Modeling.* London: Chapman & Hall/CRC.
8. Boyes, W.J., D.L. Hoffman, and S.A. Low. 1989. An Econometric Analysis of the Bank Credit Scoring Problem. *Journal of Econometrics*, 40: 3-14.

9. Blochlinger, A. and M. Leippold. 2006. Economic benefit of powerful credit scoring. *Journal of Banking and Finance*, 30: 851-873.
10. Caouette, J.B., E.I. Altman, and P. Narayanan. 1998. *Managing Credit Risk: The Next Great Financial Challenge.* New York: John Wiley & Sons Inc.
11. Chen, M., and S. Huang. 2003. Credit Scoring and Rejected Instances Reassigning Through Evolutionary Computation Techniques. *Expert Systems with Applications* 24: 433-441.
12. Desai, V.S., J.N. Crook, and G.A. Overstreet. 1996. A Comparison of Neural Networks and Linear Scoring Models in the Credit Union Environment. *European Journal of Operational Research*, 95: 24-37.
13. Durand, D. 1941. *Risk Elements in Consumer Instalment Financing,* Studies in Consumer Instalment Financing. New York: National Bureau of Economic Research.
14. Fisher, R.A. 1936. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7: 179-188.
15. Greene, W. 1998. Sample Selection in Credit-Scoring Models. *Japan and the World Economy*, 10: 299-316.
16. Guillen, M. and M. Artis. 1992. Count Data Models for a Credit Scoring System. The European Conference Series in Quantitative Economics and Econometrics on Econometrics of Duration, Count and Transition Models. Paris.
17. Hand, D.J. 1981. *Discrimination and Classification.* New York: John Wiley & Sons Inc.
18. Hand, D.J., and W.E. Henley. 1997. Statistical Classification Methods in Consumer Credit Scoring: A Review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160: 523-541.
19. Hoffmann, F., B. Baesens, C. Mues, T. Van Gestel and J. Vanthienen. 2007. Inferring descriptive and approximate fuzzy rules for credit scoring using evolutionary algorithms. *European Journal of Operational Research*, 177: 540-555.
20. Johnson, R.A., and D.W. Wichern. 2002. *Applied Multivariate Statistical Analysis.* Prentice Hall.
21. Kim, Y.S. and S.Y. Sohn. 2004. Managing Loan Customers Using Misclassification Patterns of Credit Scoring Model. *Expert Systems with Applications*, 26: 567-573.
22. Lee, T. and I. Chen. 2005. A Two-Stage Hybrid Credit Scoring Model Using Artificial Neural Networks and Multivariate Adaptive Regression Splines. *Expert Systems with Applications*, 28: 743-752.
23. Lee, T., C. Chiu, C. Lu, and I. Chen. 2002. Credit Scoring Using the Hybrid Neural Discriminant Technique. *Expert Systems with Applications*, 23: 245-254.
24. Leonard, K.J. 1995. The Development of Credit Scoring Quality Measures for Consumer Credit Application. *International Journal of Quality & Reliability Management*, 12: 79-85.
25. Lewis, E.M. 1992. *An Introduction to Credit Scoring.* California: Fair, Isaac and Co., Inc.
26. Lim, M.K. and S.Y. Sohn. 2007. Cluster-based dynamic scoring model. *Expert Systems with Applications*, 32: 427-431.
27. Long, M.S. 1973. Credit Scoring Development for Optimal Credit Extension and Management Control. College on Industrial Management, Georgia Institute of Technology. Atlanta Georgia: Purdue University.
28. Maddala, G.S. 2001. *Introduction to Econometrics.* Chichester: John Wiley & Sons Inc.
29. Mays, E. 2001. *Handbook of Credit Scoring.* Chicago: Glenlake Publishing Company, Ltd.
30. Neter, J., M.H. Kutner, W. Wasserman, and C.J. Nachtsheim. 1996. *Applied Linear Statistical Models.* Chicago: McGraw-Hill/Irwin.
31. Ong, C., J. Huang, and G. Tzeng. 2005. Building Credit Scoring Models Using Genetic Programming. *Expert Systems with Applications*, 29: 41-47.
32. Orgler, Y.E. 1971. Evaluation of Bank Consumer Loans with Credit Scoring Models. *Journal of Bank Research*, 2: 31-37.
33. Pindyck, R.S. and D.L. Rubinfeld. 1997. *Econometric Models and Economic Forecasts*. McGraw-Hill/Irwin.
34. Sarlija, N., M. Bensic, and Z. Bohacek. 2004. Multinomial Model in Consumer Credit Scoring, 10th International Conference on Operational Research. Trogir: Croatia.

35. Siddiqi, N. 2006. *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring.* New Jersey: John Wiley & Sons, Inc.
36. Steenackers, A., and M.J. Goovaerts. 1989. A Credit Scoring Model for Personal Loans. *Insurance: Mathematics and Economics*, 8: 31-34.
37. Sullivan, A.C. 1981. *Consumer Finance,* in Altman, E.I. Financial Handbook. New York: John Wiley & Sons.
38. Thomas, L.C., D.B. Edelman, and L.N. Crook. 2002. *Credit Scoring and Its Applications.* Philadelphia: Society for Industrial and Applied Mathematics.
39. Thompson, P. 1998. Bank Lending and the Environment: Policies and Opportunities. *International Journal of Bank Marketing*, 16: 243-252.
40. West, D. 2000. Neural Network Credit Scoring Models. *Computers & Operations Research*, 27: 1131-1152.
41. www.cbe.org.eg.
42. Yang, Z., Y. Wang, Y. Bai, and X. Zhang. 2004. Measuring Scorecard Performance. *Computational Science*, 3039: 900-906.

Appendix A: Statistical analysis for the proposed models

Discriminating Function for DA model:

Functions Wilks Chi-Square DF P-Value
Derived Lambda
----
1 0.543615 349.2512 12 0.0000

Discriminating Function for DA_1 model:

Functions Wilks Chi-Square DF P-Value
Derived Lambda
----
1 0.5438 349.9703 9 0.0000

Analysis of Deviance and Likelihood Ratio Tests for PA model:
Analysis of Deviance

Source Deviance Df P-Value

Model 374.5 13 0.0000
Residual 284.906 567 1.0000

Total (corr.) 659.407 580

Likelihood Ratio Tests

Factor	Chi-Square	Df	P-Value
ADD INC	0.00152616	1	0.9688
AGE	12.0717 72.313	1	0.0005
COR GUAR	11.6285	1	0.0000
EDU	2.70153	1	0.0006
HOR	72.0333	1	0.1002
LFOB	78.0624	1	0.0000
LOAN AMO	5.04102	1	0.0000
MAR STA	5.69163	1	0.0248
SALA	0.53373	1	0.0170
SEX	61.4374	1	0.4650
TELE	3.49304	1	0.0000
COMP		2	0.1744

Analysis of Deviance and Likelihood Ratio Tests for PA_1 model:
Analysis of Deviance

Source Deviance Df P-Value

Model 370.674 9 0.0000
Residual 288.732 571 1.0000

Total (corr.) 659.407 580

Likelihood Ratio Tests

Factor	Chi-Square	Df	P-Value
AGE	10.8605	1	0.0010
COR GUAR	72.5957	1	0.0000
EDU	10.7326	1	0.0011
HOR	5.60935	1	0.0179
LFOB	69.6341	1	0.0000
LOAN AMO	99.0516	1	0.0000
MAR STA	6.08719	1	0.0136
SALA	5.84293	1	0.0156
TELE	61.5081	1	0.0000

Analysis of Deviance and Likelihood Ratio Tests for LR model:
Analysis of Deviance

Source Deviance Df P-Value

Model 374.661 13 0.0000
Residual 284.746 567 1.0000

Total (corr.) 659.407 580

Likelihood Ratio Tests

Factor	Chi-Square	Df	P-Value
ADD INC	0.0171689	1	0.8958
AGE	13.4555	1	0.0002
COR GUAR	74.1195	1	0.0000
EDU	10.5227	1	0.0012
HOR	1.88777	1	0.1695
LFOB	71.4812	1	0.0000
LOAN AMO	78.5665	1	0.0000
MAR STA	4.72988	1	0.0296
SALA	5.23704	1	0.0221
SEX	0.716841	1	0.3972
TELE	62.0235	1	0.0000
COMP	3.80231	2	0.1494

Analysis of Deviance and Likelihood Ratio Tests for LR_1 model:
Analysis of Deviance

Source Deviance Df P-Value

Model 370.372 9 0.0000
Residual 289.035 571 1.0000

Total (corr.) 659.407 580

Likelihood Ratio Tests

Factor	Chi-Square	Df	P-Value
AGE	12.3538	1	0.0004
COR GUAR	73.6767	1	0.0000
EDU	9.75523	1	0.0018
HOR	4.86088	1	0.0275
LFOB	68.7425	1	0.0000
LOAN AMO	99.7909	1	0.0000
MAR STA	6.12316	1	0.0133
SALA	5.35199	1	0.0207
TELE	61.8505	1	0.0000

Appendix B: Average correct classification rates for PA, PA_1, LR, LR_1

Cut-off / Model	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75
PA	0.8640	0.8589	0.8675	0.8744	0.8778	0.8795	0.8898	0.8933	0.8744	0.8640
PA_1	0.8640	0.8589	0.8675	0.8761	0.8726	0.8795	0.8881	0.8795	0.8675	0.8623
LR	0.8606	0.8675	0.8709	0.8761	0.8830	0.8881	0.8985	0.8916	0.8761	0.8623
LR_1	0.8640	0.8675	0.8761	0.8744	0.8795	0.8847	0.8916	0.8761	0.8675	0.8640

Numbers in cells refer to the average correct classification rates under the different cut-offs. The 0.50 standard cut-off rates and the highest rates per model are highlighted.