A two-stage dynamic credit scoring model, based on customers' profile and time horizon

Received (in revised form): 17th February, 2008

Maria Mavri*

is Lecturer of Quantitative Methods in the School of Business, Department of Business Administration, University of the Aegean, Chios, Greece. She holds a first degree in Mathematics from the University of Athens, an MSc degree in Decision Sciences from Athens University of Business and Economics and a PhD in Operational Research from the same university. Her scientific interests apply in the field of Operational Research; Data Analysis, Linear and nonlinear Programming, Optimisation, Survival Analysis, Banking Management and Customer Relationship Management. She has presented her research work in international conferences and she has published research papers in scientific journals such as *Omega*, *International Transactions in Operational Research* and *International Journal of Consumer Studies*.

Vassilis Angelis

is Professor of Quantitative Methods in the School of Business, Department of Business Administration, University of the Aegean, Chios, Greece. He holds a first degree in Mathematics from the University of Athens, an MSc degree in Industrial Mathematics and Statistics from the University of Aston in Birmingham and a PhD in Control Theory from the University of Warwick. He teaches graduate and undergraduate courses in Quantitative Methods. His research areas of interest include Sustainable Regional Development, Location Analysis, Business Process Management, Performance and Customer Satisfaction in the Services Sector. He has published extensively in journals such as *Studies in Location Analysis*, *The Journal of Management Science and Regional Development*, *Studies in Urban and Regional Planning*, *International Journal of Applied Systemic Studies*, *Business Process Management* and *The Journal of Financial Services Marketing*.

George Ioannou

is an Associate Professor at the Athens University of Economics and Business, the Acting Director of the International MBA Program, and Head of the Operations Management & ERP Systems Center. He received his diploma in Mechanical Engineering from NTUA (Athens), his MSc/DIC in Industrial Robotics from Imperial College and his PhD in Mechanical Engineering from the University of Maryland. His research concentrates on business systems, merging operations research with information technology. He is the recipient of the Microsoft Excellence in Education Award, chairs the Research and Development Committee of the Technical Chamber of Greece and is an Editorial Board member of *Production Planning & Control*.

Eleni Gaki

is a member of Special Scientific and Teaching Staff in the School of Business, Department of Business Administration, University of the Aegean, Chios, Greece. She has earned her PhD in Economic and Regional Analysis from the University of the Aegean. She teaches graduate and undergraduate courses in Quantitative Methods. Her research areas of interest include Transport/ICT and Regional Development, Isolated Regions' Development. She has presented several papers in international conferences and has published papers in *Studies in Urban and Regional Planning*.

Iason Koufodontis

is adjunct Instructor in the School of Business, University of the Aegean, Chios, Greece. He has earned an MSc in Tourism Management, Policy and Planning from the University of the Aegean and his PhD in ICT and Innovative Organisations from the same university. He teaches graduate courses in ICT. His research areas of interest include Information and Communication Technologies and Regional Development, Innovative and Network Organisations, Destination Management and use of ICT in Tourism. He has presented several papers in international conferences.

Abstract As credit card usage has expanded rapidly worldwide, credit scoring has become a very important task for banks, which can benefit from reducing possible risks of default. Credit scoring models help decision makers to decide whether to issue a credit card to a new

*Correspondence: Quantitative Methods Laboratory, Department of Business Administration, University of the Aegean, Michalon 8, Chios 82100, Greece.
Tel: +30 2270 35184;
Fax: +30 22710 35189;
e-mail: m.mavri@ba.aegean.gr

applicant on the basis of both financial and nonfinancial criteria. The scope of the current study is to develop a dynamic scoring model that (a) estimates the credit performance of an applicant using generalised linear models and (b) accommodates the changes of a borrower's characteristics after the issuance of the credit card and forecasts the time of default using survival analysis.

Journal of Financial Services Marketing (2008) **13,** 17–27. doi:10.1057/fsm.2008.2

Keywords credit scoring model, logistic regression, survival analysis

INTRODUCTION

Socio-economic changes on a global level, coupled with technological advancements, have changed the traditional banking environment. Over the last three decades the cash-oriented society has been transformed into a 'plastic money' society. The difference lies in the fact that more and more business is now transacted in a computer-mediated environment. Transactions, bill payments, purchasing, reservations (hotels, travel or cinema tickets) and other similar activities tend to be carried out electronically. The demand for credit card issuance has been augmented dramatically and, as a result, the bank industry has been prompt to develop credit scoring models in order to estimate quickly and accurately the credit performance of each applicant and to decide whether to issue a credit card. Scoring models have become of primary importance in the financial environment.

Credit scoring has gained attention as the credit industry benefits from timely decisions, reduction of possible risks, improved cash flow and proper credit collections. Thomas[1] identified two types of decisions that banks who lend to consumers have to make. First, they should grant credit to a new applicant, and secondly they have to continue and extend their cooperation with existing customers. Techniques that help with these decisions are called credit and behavioural scoring models. The main objective for both types of scoring models is to classify customers into groups. Applied to bank databases, classification analysis for credit scoring is used to categorise a new applicant as either 'accepted' or 'rejected' with respect to their features such as age, income and martial status.[2] Classification analysis for behavioural scoring models is used to describe the behaviour of existing customers and to predict future purchasing behaviour status.[3]

Credit scoring models and their classification techniques are under examination in this study. The accuracy of their estimations over a period of time is an emerging issue in the field of financial scoring. The majority of these models assign credit applicants to either a 'good credit' group, which is likely to repay a financial obligation, or a 'bad credit' group, with a high probability of defaulting on the financial obligation and hence their application should be denied,[4] but they do not consider information related to the timing of the events.

The purpose of this paper is to develop a two-stage dynamic Credit Scoring Model in order to estimate (a) the risk level of a credit card applicant and (b) the probability of default during a pre-specified period of time, the length of which is determined by the bank's management team. The proposed model begins by evaluating the financial credibility of all applicants, and continues by predicting future repayment behaviour for those who have been characterised as creditworthy customers. By taking into account demographic and socio-economic data, the model determines factors that significantly improve an applicant's risk level. The dimension of time that has been taken into account in those estimations is a basic difference of the proposed model as compared to other similar models that can be found in the literature.

The remainder of this paper is organised as follows: The next section gives a brief overview of the existing literature. The subsequent section presents the proposed model, while the penultimate section looks at the model's application on a real case and discusses the results. Finally, the last section summarises the conclusions and suggests areas for further research.

LITERATURE REVIEW

Credit scoring modeling has become a core component in risk management systems at banks and financial institutions. In fact, banks are prompt to develop or buy such models in order to make the whole procedure of evaluating credit card applications faster, easier and more accurate. The objective of credit scoring models is to decide whether or not to grant credit to an applicant. Therefore, credit scoring models basically belong to the field of classification problems.[5]

Common methodologies take a sample of existing customers and classify them into good or bad depending on their repayment performance over a fixed period.[6–8] The most commonly used credit measurement methodologies are (a) discriminant analysis,[9–11] and (b) logistic regression analysis.[12,13] Both have the merits of being conceptually straightforward.

Discriminant analysis is a predictive model of group membership based on observed characteristics of each group member. It was first proposed by Fisher[14] and looks for the best linear combination of the predictor variables in order to classify the factors under examination into two or more groups.[1,5] Based on the above criterion, an observation is classified into a group on the basis of a set of predictor variables. However, linear discriminant analysis has often been criticised because of the following strict prerequisites it sets: (1) The nature of credit data must be categorical; (2) The data must be independent and normally distributed; (3) The covariance matrices of the good and bad credit groups are unlikely to be equal. Moreover, the obtained results do not provide any estimate of the associated risk.

Linear regression analysis is widely used in order to relate the probability of a binary outcome with a set of explanatory variables. Logistic regression models can estimate the probability of an applicant's credit status and give a better understanding of the financial risk distribution than the discriminant analysis approach. More specifically, they determine the conditional probability of an applicant to repay a financial obligation, given the values of the independent variables of the credit applicant.[4] Furthermore, those models do not necessarily set the strict application prerequisites of the discriminant analysis approach.

In cases where the relationship between the dependent and independent variables is not linear and hence the two previous models are not applicable, neural networks may be used.[15–17] Neural networks process information through the interactions of a large number of processing units and their connections to external inputs. They are found to be increasingly useful in modelling nonstationary processes due to their outstanding generalisation capability.[18] However, they have been criticised for the length of the process in obtaining the optimal network's topology and the difficulty in identifying the importance of potential input variables.

The performance of a credit scoring model using the hybrid modelling procedure with artificial neural networks and multivariate adaptive regression splines (MARS) is explored by Lee *et al.*[19] They used MARS in order to identify the significant variables, which may then serve as input nodes to the neural networks model.

Regression spline approach models the mean outcome as a piecewise polynomial, such as a piecewise continuous linear function or a piecewise cubic function with continuous derivative of a predictor variable. A piecewise polynomial function $f(x)$ can be obtained by dividing the range of each predictor variable into one or more intervals and representing the function f by a separate polynomial in each interval.

Lee *et al.*[19] developed another credit scoring model by using two data mining techniques: classification and regression tree (CART) and multivariate adaptive regression splines. Both these models give better results as compared to classical credit scoring models. However, they have certain limitations related to interpretative difficulties.

All the above models are useful in data classification, but they do not consider information related to the timing of the events and cannot handle time-dependent covariates. Hyun *et al.*[6] looked at the extension of the survival analysis model to analyse personal credit risk. In their study, they used a survival analysis approach in order to determine the time when 'good' customers become 'bad' as well as the default probability of credit applicants or existing customers.

Summarizing, we can say that the existing literature focuses on credit risk models based on binary classification approaches. Most of these models classify an applicant according to their characteristics as shown in their application form. However, the change of the characteristics over time is not an issue under consideration. Although recent researchers look at scoring approaches using data mining techniques, the length of the observation period to predict a borrowers' credibility is a point that needs further attention.

THE PROPOSED MODEL

Having completed the brief literature review, we will now go on to present the proposed model. Since logistic regression analysis and survival analysis are the basic statistical techniques used in the process of building our model, we will start by presenting their key points as seen from the viewpoint of their utilization in our model. Then we will go on to describe the problem on which our model will be applied and outline its solution methodology. Finally, all the solution steps and the statistical techniques used for their implementation will be integrated into the proposed model, which is formulated accordingly.

Statistical techniques

Logistic regression

Logistic regression analysis is widely used in order to relate the probability p_j of a binary outcome with a set of explanatory variables x_{ij}. The probability p_j is expressed as follows:

$$p_j = \frac{exp(\beta_0 + \beta_1 x_{1j} + \ldots + \beta_k x_{kj})}{1 + exp(\beta_0 + \beta_1 x_{1j} + \ldots + \beta_k x_{kj})} \qquad (1)$$

where x_{ij} are the explanatory variables, β_I are the regression coefficients of independent variables and β_0 is the constant term.

Survival analysis

Survival analysis is a collection of statistical methods for data analysis used when the outcome variable of interest is the time from the beginning of an individual's observation until an event occurs. We usually refer to this event as 'failure' and to the time up to this event as survival time (T), since it gives an indication of how long an individual has survived over some follow-up period. At this point, it should be noted that the time to failure may be fully observed for some individuals, but only partly for others who are somehow 'censored' (ie lost) at some point in time ('censoring' time) in the observation period.

The probability that an individual's survival time is longer than a given value t, is expressed by survival function $S(t)$ which is defined as follows:

$$S(t) = P(T > t) = 1 - F(t) \quad (2)$$

where

$$F(t) = P\{T \leqslant t\} \quad (3)$$

is the cumulative distribution function of survival time.

There are several ways of presenting survival functions in a curve form. In our case since time horizon is specified, the Kaplan–Mayer method is used.

Problem description

Consider a financial institution or a bank, where many application forms for credit card issuance have just been received. The credit card department must estimate quickly and accurately the credit risk level of each applicant and decide whether or not to issue a credit card. Information about each individual's banking activity (debts, number and types of cards they hold, etc) as well as demographic and market data are provided within the application, submitted to the bank and can be utilised in order to forecast the potential of each individual's application. Our objectives in this work are to

- identify the factors on the basis of which an applicant is classified as suitable for holding a credit card,
- estimate the credit performance of each applicant and classify applicants into two groups, 'good' and 'bad',
- estimate the time of the default probability for each 'good' applicant and
- identify the factors that determine the probability of default within a period of time.

The steps taken towards achieving those objectives are described below.

Identification of explanatory variables

In developing a credit scoring model, controllable and noncontrollable factors, which influence the risk level of an applicant, must be considered. Those factors are expressed by the respective variables x_{ij}, where i and j are indices referring to the specific factor and applicant, respectively. The values of all explanatory variables for a specific applicant are included in her application form. The 11 independent variables that have been taken into account in this work are presented in Table 1.

It should be noted that variables x_{2j} and x_{5j}, to which we will be referring in the course of this paper, are defined as follows

$$x_{2j} = \begin{cases} 1 & \text{age} <= 27 \\ 2 & 28 <= \text{age} <= 40 \\ 3 & 41 <= \text{age} <= 50 \\ 4 & \text{age} > 50 \end{cases}$$

$$\text{and} \quad x_{5j} \begin{cases} 1 & \text{income} <= 800 \\ 2 & 801 < \text{income} <= 1700 \\ 3 & \text{income} > 1700 \end{cases}$$

Estimation of an applicant's risk level

The risk level of an applicant is defined by a set of explanatory variables and is measured by the logistic regression equation. In other words, the probability that the bank will issue a credit card for the j-th applicant with n-specific characteristics as expressed by the explanatory variables x_{ij}, is given by p_i as defined in the section Logistic Regression. (1). By applying logistic regression analysis, the model identifies the statistically significant variables and uses them to calculate the probability p_i for every applicant, thus classifying applicants into two groups: credible and noncredible. If this probability is equal to or greater than 0.5, then the respective application is accepted. The value of 0.5, as a cut-off value, is widely used in the literature.[20,21] Through LRA, factors that can

Table 1 Explanatory variables

Type	*Notation*	*Name*	*Scale*
Noncontrollable (demographic data)	x_{1j}	Gender	Binary: 1=male, 0=female
	x_{2j}	Áge	Ordinal
	x_{3j}	Education	Ordinal
	x_{4j}	Marital status	Binary: 1=married, 0=otherwise
Controllable factors (economic data)	x_{5j}	Monthly income	Ordinal
	x_{6j}	Period of time in the same work	Binary: 1=a year or more, 0=otherwise
	x_{7j}	Financial credibility	Binary: 1=bankruptcy, 0=otherwise
	x_{8j}	Own property	Binary: 1=yes, 0=no
	x_{9j}	Holder of other credit cards	Binary: 1=yes, 0=no
	x_{10j}	Banking activity	Binary: 1=yes, 0=no
	x_{11j}	Preferred credibility limit	Continuous

significantly improve the probability of credit card acceptance are also identified.

Estimation of a 'good' applicant's probability of default

Apart from estimating an applicant's risk level, the bank is also interested in calculating the probability of default for all applicants who have been classified as 'Credible'. By applying survival analysis to the group of credible applicants, the model estimates this probability as well as the failure time of each 'good' applicant over a specified period of time.

Model presentation

The proposed dynamic credit scoring model consists of two stages: A and B. Stage A classifies all potential credit card-holders into two groups, credible and noncredible, while Stage B focuses on the credible applicants and calculates the probability of default and the failure time. The steps taken in every stage are described below.

Stage A

Step A1: *Input of the:*

- *total number of applicants (m)*
- *number of explanatory variables (n) and their values*

Step A2: *Application of logistic regression on controllable and noncontrollable explanatory variables*

Step A3: *Identification of the statistically significant variables*

Step A4: *Estimation of all applicants' risk level*

Step A5: *Classifications of applicants into Group 1 'credible applicants' and Group 2 'noncredible applicants'*

Stage B

Step B1: *Input of the:*

- *number of Group 1 credible applicants (l),*
- *number of statistically significant explanatory variables ($k<n$) and their values (we use only the variables identified as statistically significant in Stage A).*

Step B2: *Application of survival analysis (Kaplan–Mayer)*

Step B3: *Estimation of the time until the event will occur, that is, a credible applicant will default*

Step B4: *Elaboration on the differences among credible applicants*

Figure 1 gives a schematic overview of the two-stage dynamic credit scoring model presented above.

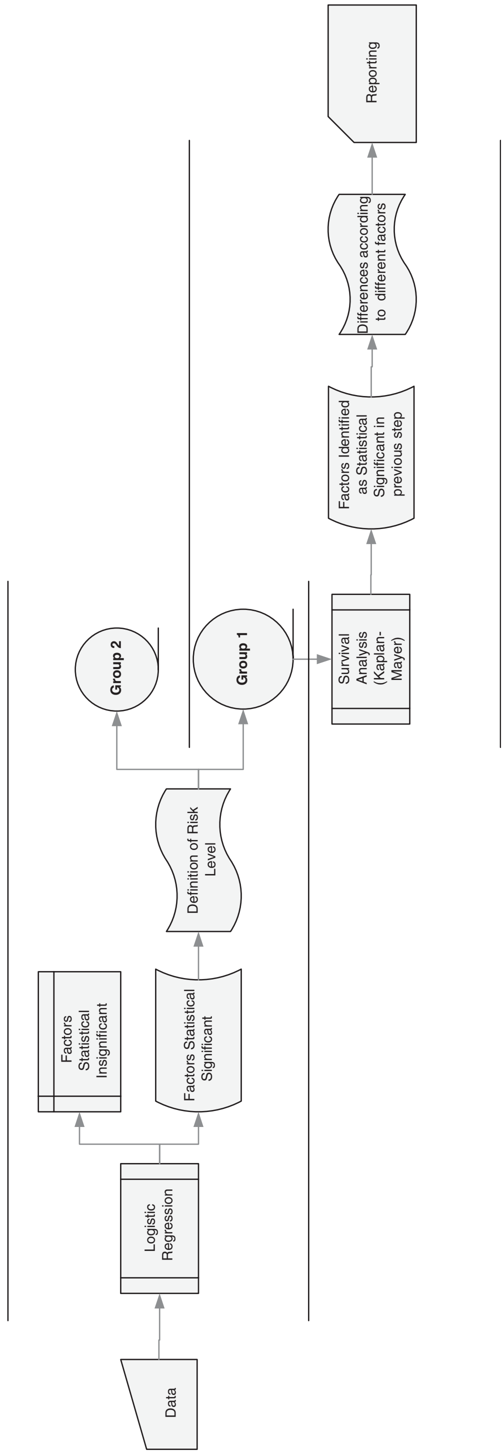

Figure 1 Credit scoring model overview

MODEL APPLICATION

The proposed two-stage dynamic credit scoring model has been applied on data drawn from a leading European bank ($n = 350$). Applicants were asked to fill out a form for credit card issuance, containing information such as property, banking activity, preferred credit card limit, age, gender, education, profession, current work status, monthly income and marital status.

The application of Stage A of the model to our data shows that the risk level of an applicant is determined by monthly income, age, own property and financial credibility. Since the level of significance of factors like education and holding other credit cards was found to be low, these factors have been parameterized in the constant term of the logistic equation. It should be noted that the two variables of monthly income and age were found to be statistically significant for applicants belonging to two age groups 28–40 years and 41–50 years, with monthly income of €801–€1,700 or above €1,700.[22] Stage A concludes with the applicants' database segmentation into Group 1 credible applicants and Group 2 noncredible applicants. The model estimates that Group 1 contains 46 per cent of the applicants (ie 158), thus indicating that the bank will issue a credit card only to them. Having defined Group 1 and Group 2, we will proceed to Stage B and continue our analysis with the credible applicants ($l = 158$).

The application of Stage B of the model to the credible applicants gives an estimate of their probability of default over a time horizon of 25 years as set by the bank's management. The model shows that the mean default time is estimated to be 15.4 years. The rather long survival time is explained by the mentality and the culture of Greek bank customers, the vast majority of which remain, under normal conditions, loyal to their bank.

It should be noted that in estimating this mean time, only factors identified as

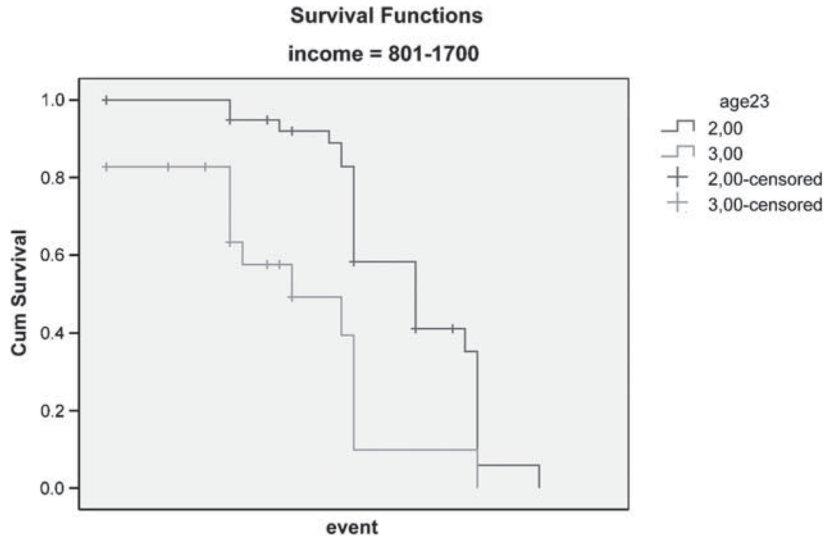

Figure 2 Survival function (Income between 801 and 1,700€)

statistically significant in the previous stage have been used. Hence, characteristics such as financial condition and property, which are common for all applicants of Group 1, have not been taken into account.

Figures 2 and 3 present the differences in survival time among the applicants due to differences in income and age, respectively.

Figure 2 presents the survival functions for credible applicants who have a monthly income between €801 and €1,700 and belong to two different age categories: category (a) 28–40 years old and category (b) 41–50 years old. Category (a) appears to have consistently better survival prognosis than category (b). The mean time until the event of default will occur is estimated to be 17 and 12 years, respectively.

Figure 3 presents the survival functions for credible applicants who have a monthly income of more than €1,700 and belong to the age categories (a) and (b), as defined above. By comparing the two plots, we conclude again that the prognosis for category (a) is better than category (b). However, the two curves are very similar during the first five years, but quite different after ten years. The mean survival time is approximately 17.5 and 14 years, respectively. The longer default time in the case of younger applicants is explained by the fact that they have a potentially longer cooperation time with the bank.

At this point it is also interesting to observe the two Receiver Operation Characteristic (ROC) curves provided by the two stages of the model, using logistic regression analysis and survival analysis, respectively (Figures 4 and 5).

Both the above curves are plots of the true positive rate (sensitivity) against the false positive rate (1-specificity) for the different possible cut-off points of a classifier. It is noted that the true positive rate (sensitivity) expresses the percentage of credible applicants correctly classified by the model, whereas the false positive rate expresses the percentage of credible applicants wrongly classified.

By looking at those curves, we can also conclude that

- in both of them any increase in sensitivity is accompanied by a decrease in specificity and
- both of them are located left to the diagonal, implying the correctness of the classification criterion in both LRA (credible and noncredible applicants) and survival analysis (switching behaviour or not).

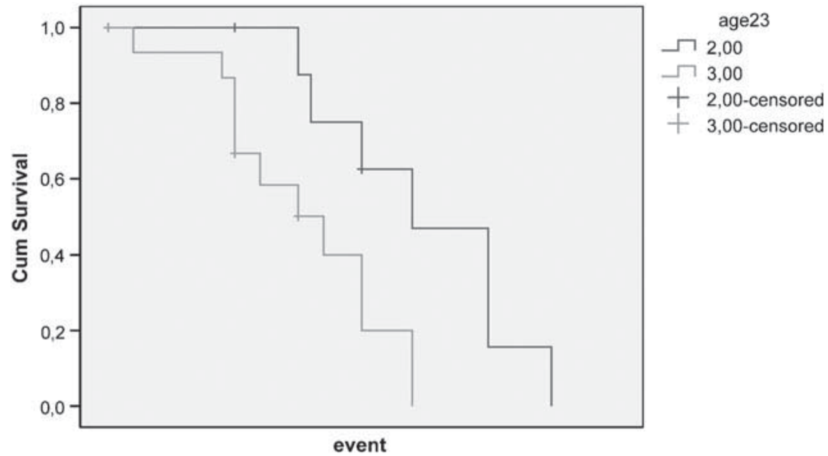

Figure 3 Survival function (Income more than 1,700€)

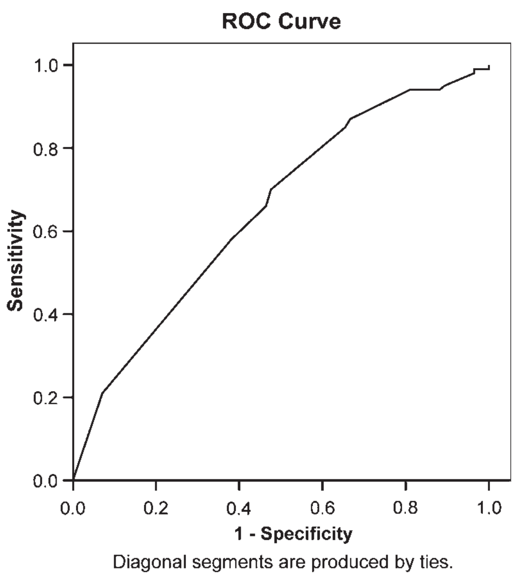

Figure 4 ROC curve LRA

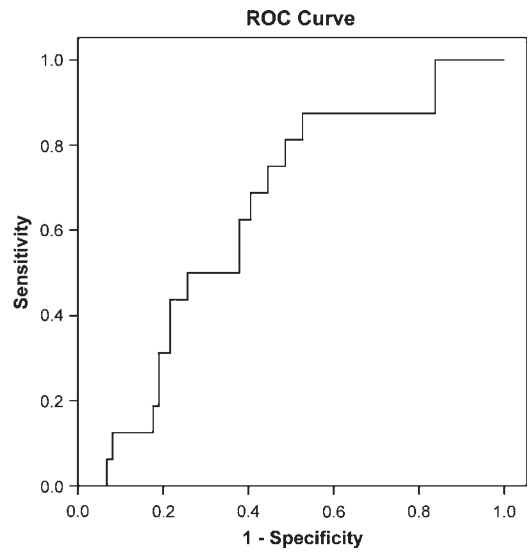

Figure 5 ROC curve SA

We have so far in this work used logistic regression analysis for classifying our data into two groups of credible and noncredible applicants. An alternative method for grouping data is the use of discriminant analysis as mentioned in the Literature Review section. Table 2 presents an overall measure of the classification correctness for both logistic regression analysis and discriminant analysis. This measure expresses the average percentage of the credible and noncredible applicants classified correctly by the model.

Table 2 Overall correctness of classification

	Logistic regression analysis (%)	*Discriminant analysis (%)*
Average percentage of credible and non-credible applicants correctly classified	71.87	60.5

As we can see, the logistic regression analysis, which has been used in our model, gives better results. This is consistent with

Table 3 Correctness of noncredible applicants' classification

	Logistic regression analysis (%)	*Discriminant analysis (%)*
Percentage of non-credible applicants correctly classified	71.7	59.8

similar findings reported in the literature (Desai *et al.*,[9] Harrell and Lee[23]).

Banks' experience has shown that the cost of issuing a credit card to a defaulter is greater than rejecting a good applicant. Hence, the percentage of noncredible applicants correctly identified, is another important measure of the model's effectiveness. Table 3 presents the percentage of noncredible applicants correctly classified by the two different techniques.

As in the previous case, the logistic regression analysis gives better results.

CONCLUSIONS — SUGGESTIONS FOR FURTHER RESEARCH

In this paper, we have presented a model assessing the credibility of potential credit card-holders, thus classifying them into credible and noncredible and then estimating the probability of default and the survival time of those applicants who have been classified as credible. A number of demographic and financial factors have been used for those estimations. The dimension of time that has also been taken into account in those estimations is a basic difference of the proposed model as compared to other similar models that can be found in the literature.

By applying the proposed model on a sample of 350 applications for credit card issuance, it has been found that factors such as the applicants' monthly income, age, owned property and financial credibility up-to-date determine the risk level. Furthermore, the survival time of those applicants who have been classified as credible is on average 15.1 years and depends on monthly income and age. More specifically, it has been found that irrespective of differences in monthly income the survival time is longer for people belonging to the age group 28–40 years, than people belonging to the age group 41–50 years. It must be, however, noted that for applicants of the latter age group any increase in their monthly income improves the survival time.

Summarizing all the above, we can say that the proposed model may prove a very valuable tool for a bank's management and more specifically for the Credit Card Approval Department, since it will provide useful information for the applicants' credibility and survival time. Furthermore, if we take into account that the cost of seeking new customers is extremely high, we realise the critical importance of the model's ability to estimate survival time for existing customers and consequently our ability of taking measures for its extension.

Elaboration on the changes in the characteristics of credible customers over time may be a field for further research. Furthermore, the adjustment of the model so as to include stochastic variables will make it more reliable and useful for the bank's management. The reason is that, apart from survival time, the model will be able to identify at the same time, the characteristics of a given applicant that may probably lead at a future time to her transformation from credible to noncredible customer.

Another area of further research would be to transform the proposed Credit Scoring Model into a Decision Support System for the management of banks going through a transition period in the new economy. To do this, a number of steps should be taken, including the automatic data collection from local management information systems, the automatic survival analysis on results drawn from logistic regression analysis and the comprehensive user interface for conveying the results in the proper format for management.

References

1 Thomas, L. C. (2000) 'A survey of credit and behavioural scoring: Forecasting financial risk of lending to consumers', *International Journal of Forecasting*, Vol. 16, pp. 149–172.

2 Chen, M. C. and Huang, S. H. (2003) 'Credit scoring and rejected instances reassigning through evolutionary computation techniques', *Expert Systems with Applications*, Vol. 24, pp. 433–441.

3 Hsieh, N. -C. (2004) 'An integrated data mining and behavioral scoring model for analyzing bank customers', *Expert Systems with Applications*, Vol. 27, pp. 623–633.

4 Lee, T. -S. and Chen, I. -F. (2005) 'A two stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines', *Expert Systems with Applications*, Vol. 28, pp. 743–752.

5 Johnson, R. A. and Wichern, D. W. (2002) 'Applied Multivariate Statistical Analysis', 5th edn, Prentice-Hall, Upper Saddle River, NJ.

6 Hyun, J. H., Roh, H. T. and Han, I. (2005) 'Prognostic personal credit risk model considering censored information', *Expert Systems with Applications*, Vol. 28, pp. 753–762.

7 Beasens B., Van Gestel, T., Vianene, S., Stephanova, M., Suykens, J. and Vanthienene, J. (2003) 'Benchmarking state of the art classification algorithms for credit scoring', *Journal of the Operational Research Society*, Vol. 54, No. 6, pp. 627–635.

8 Stephanova, M. and Thomas, L. (2002) 'Survival analysis methods for personal loan data', *Operations Research*, Vol. 50, No. 2, pp. 277–289.

9 Desai, V. S., Crook, J. N. and Overstreet Jr., G. A. (1996) 'A comparison of neural networks and linear scoring models in credit union environment', *European Journal of Operational Research*, Vol. 95, pp. 24–37.

10 Bardos, M. (1998) 'Detecting the risk of company failure at the banque de France', *Journal of Banking and Finance*, Vol. 22, pp. 1405–1419.

11 Kim, J. C., Kim, D. H., Kim, J. J., Ye, J. S. and Lee, H. S. (2000) 'Segmenting the Korean housing market using multiple discriminant analysis', *Construction Management and Economics*, Vol. 18, pp. 45–54.

12 Hierr, L. and Frei, F. (2002) 'Do better customers utilize electronic distribution channels? The case of PC banking', *Management Science*, Vol. 48, No. 6, pp. 732–748.

13 Daniel, E. (1999) 'Provision of electronic banking in the UK and the Republic of Ireland', *International Journal of Bank Marketing*, Vol. 38, No. 1, pp. 8–14.

14 Fisher, R. A. (1936) 'The use of multiple measurements in taxonomic problems', *Annals of Eugenics*, Vol. 7, pp. 179–188.

15 West, D. (2000) 'Neural network credit scoring models', *Computers and Operations Research*, Vol. 27, pp. 1131–1152.

16 Lee, T. -S., Chiu, C. C., Lu, C. -J. and Chen, I. -F. (2002) 'Credit scoring using the hybrid neural discriminant technique', *Expert Systems with Applications*, Vol. 23, pp. 225–234.

17 Piramuthu, S. (1999) 'Financial credit risk evaluation with neurofuzzy systems', *European Journal of Operational Research*, Vol. 112, pp. 310–321.

18 Stern, H. S. (1996) 'Neural networks in applied statistics', *Technometrics*, Vol. 38, No. 3, pp. 205–216.

19 Lee, T. -S., Chiu, C. C., Chou, Y. -C. and Lu, C. -J. (2006) 'Mining the customer credit using classification and regression tree and mulrtivariate adaptive regression splines', *Computational Statistics and Data Analysis*, Vol. 50, pp. 1113–1130.

20 Dobson, A. (2002) 'An Introduction to Generalized Linear Models', 2nd edn, Chapman & Hall, London.

21 Collett, D. (1991) 'Modeling Binary Data', 1st edn, Chapman & Hall, London.

22 Mavri, M. and Ioannou, G. (2004) 'An empirical study for credit card approvals in the Greek banking sector', *Operational Research, An International Journal*, Vol. 4, No. 1, pp. 29–44.

23 Harrell, F.E. and Lee, K.L. (1985) 'A comparison of the discrimination of discriminant analysis and logistic regression', in Se, P.K. (ed), 'Biostatistics: Statistics in Biomedical Public Health and Environmental Sciences', Amsterdam.