



University of
Salford
MANCHESTER

Credit scoring, statistical techniques and evaluation criteria: A review of the literature

Abdou, HAH and Pointon, J

[/http://dx.doi.org/10.1002/isaf.325](http://dx.doi.org/10.1002/isaf.325)

Title	Credit scoring, statistical techniques and evaluation criteria: A review of the literature
Authors	Abdou, HAH and Pointon, J
Type	Article
URL	This version is available at: http://usir.salford.ac.uk/16518/
Published Date	2011

USIR is a digital collection of the research output of the University of Salford. Where copyright permits, full text material held in the repository is made freely available online and can be read, downloaded and copied for non-commercial private study or research purposes. Please check the manuscript for any further copyright restrictions.

For more information, including our policy and submission procedure, please contact the Repository Team at: usir@salford.ac.uk.

Full reference: Abdou, H. & Pointon, J. (2011) 'Credit scoring, statistical techniques and evaluation criteria: a review of the literature ', *Intelligent Systems in Accounting, Finance & Management*, 18 (2-3), pp. 59-88.

Credit Scoring, Statistical Techniques and Evaluation Criteria: A Review of the Literature

Hussein A. Abdou*

Salford Business School, University of Salford, Salford, Greater Manchester, M5 4WT, UK

John Pointon

Plymouth School of Management, University of Plymouth, Plymouth, Devon, PL4 8AA, UK

Summary

Credit scoring has been regarded as a core appraisal tool of different institutions during the last few decades, and has been widely investigated in different areas, such as finance and accounting. Different scoring techniques are being used in areas of classification and prediction, where statistical techniques have conventionally been used. Both sophisticated and traditional techniques, as well as performance evaluation criteria are investigated in the literature. The principal aim of this paper is to carry out a comprehensive review of 214 articles/books/theses that involve credit scoring applications in various areas, in general, but primarily in finance and banking, in particular. This paper also aims to investigate how credit scoring has developed in importance, and to identify the key determinants in the construction of a scoring model, by means of a widespread review of different statistical techniques and performance evaluation criteria. Our review of literature revealed that there is no overall best statistical technique used in building scoring models and the best technique for all circumstances does not yet exist. Also, the applications of the scoring methodologies have been widely extended to include different areas, and this subsequently can help decision makers, particularly in banking, to predict their clients' behaviour. Finally, this paper also suggests a number of directions for future research.

Keywords: Credit Scoring; Classification Techniques; Performance Evaluation Criteria; Literature Review.

*** Correspondence Author**

Dr. Hussein Abdou
Senior Lecturer in Finance & Banking
Salford Business School
University of Salford
Salford
Greater Manchester
M5 4WT
UK
Tel.: +44 1612 953001
Fax: +44 1612 955022
Email: h.abdou@salford.ac.uk

1. Introduction

The phenomenon of borrowing and lending has a long history associated with human behaviour (Thomas et al., 2002). Therefore, credit is perhaps a phenomenon as old as trade and commerce. Despite the very long history of credit back to around 2000 BC or earlier, the history of credit scoring is very short, beginning only about six decades ago. Information collected by banks and/or financial institutions of a credit applicant is used to develop a numerical score for each applicant (Thomas et al., 2002; Hand & Jacka, 1998; Lewis, 1992). Recently, credit scoring techniques have been expanded to include more applications in different fields. Moreover, the idea of reducing the probability of a customer defaulting, which predicts customer risk, is a new role for credit scoring, which can support and help maximize the expected profit from that customer for financial institutions, especially banks. By the start of the 21st century, the use of credit scoring had expanded more and more, especially with the tremendous technologies created, introducing more advanced techniques and evaluation criteria, such as GINI and area under the ROC curve¹. Besides, the high capabilities of computing technology make the use of credit scoring much easier than before.

Consequently the history of credit scoring is short, and the literature is very limited. Books that have been introduced are limited (see, for example, Lewis, 1992; Hand & Jacka, 1998; Mays, 2001, 2004; Cramer, 2004; Siddiqi, 2006; Anderson, 2007); textbooks looking at classification problems are also limited (Hand, 1981, 1997), whilst, in recent years, a number of international journal articles have discussed different credit scoring techniques in different fields (see, for example, Desai et al., 1996; Leonard, 1996; Thomas, 1998; West, 2000; Baesens et al., 2003; Lee & Chen, 2005; Lensberg et al., 2006; Banasik & Crook, 2007; Huang et al., 2007; Paliwal & Kumar, 2009).

1.1. Credit scoring definitions

Credit evaluation is one of the most crucial processes in banks' credit management decisions. This process includes collecting, analysing and classifying different credit elements and variables to assess the credit decisions. The quality of bank loans is the key determinant of competition, survival and profitability. One of the most important kits, to classify a bank's customers, as a part of the credit evaluation process to reduce the current and the expected risk of a customer being bad credit, is credit scoring. Hand & Jacka, (1998, p. 106) stated that "the process (by financial institutions) of modelling creditworthiness is referred to as *credit scoring*". It is also useful to provide further definitions of credit scoring.

¹ Both are tools to evaluate the predictive performance of different scoring models. The ROC curve is a graphical plot of the sensitivity versus 1-specificity for a dichotomous classifier (to discriminate between two classes); whilst, the GINI coefficient is a measure of the inequality of a distribution, and summarizes the predictive performance over all cut-off score values, more details are provided in section 2.5.

Anderson (2007) suggested that to define credit scoring, the term should be broken down into two components, credit and scoring. Firstly, simply the word 'credit' means "buy now, pay later". It is derived from the Latin word 'credo', which means 'I believe' or 'I trust in'. Secondly, the word 'scoring' refers to "the use of a numerical tool to rank order cases according to some real or perceived quality in order to discriminate between them, and ensure objective and consistent decisions". Therefore, scores might be presented as "numbers" to represent a single quality, or "grades" which may be presented as "letters" or "labels" to represent one or more qualities (Anderson, 2007, pp. 3-5). Consequently, credit scoring can be simply defined as "the use of statistical models to transform relevant data into numerical measures that guide credit decisions. It is the industrialisation of trust; a logical future development of the subjective credit ratings (see for example, Beynon, 2005) first provided by nineteenth century credit bureaux, that has been driven by a need for objective, fast and consistent decisions, and made possible by advances in technology" (Anderson, 2007, p. 6). Furthermore, "Credit scoring is the use of statistical models to determine the likelihood that a prospective borrower will default on a loan. Credit scoring models are widely used to evaluate business, real estate, and consumer loans" (Gup & Kolari, 2005, p. 508). Also, "Credit scoring is the set of decision models and their underlying techniques that aid lenders in the granting of consumer credit. These techniques 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" (Thomas et al., 2002, p. 1).

Credit scoring models (see, for example: Lewis, 1992; Bailey, 2001; Mays, 2001; Malhotra & Malhotra, 2003; Thomas et al., 2004; Sidique, 2006; Chuang & Lin, 2009; Sustersic et al, 2009) are some of the most successful applications of research modelling in finance and banking, as reflected in the number of scoring analysts in the industry, which is continually increasing. "However, credit scoring has been (vital) in allowing the phenomenal growth in consumer credit over the last five decades. Without (credit scoring techniques, as) an accurate and automatically operated risk assessment tool, lenders of consumer credit could not have expanded their loan (effectively)" (Thomas et al, 2002, p. xiii).

1.2. Judgemental systems versus credit scoring systems

The overall idea of credit evaluation is to compare the features or the characteristics of a customer with other earlier period customers, whose loans they have already paid back. If a customer's characteristics are adequately similar to those, who have been granted credit, and have consequently defaulted, the application will normally be rejected. If the customer's features are satisfactorily like those, who have not defaulted, the application will normally be granted. Generally, two techniques can be used: "Loan officer's subjective assessment and credit scoring" (Crook, 1996).

Sullivan (1981) and Bailey (2004) argue that in a judgemental technique evaluation, each credit application includes information contained within it, to be evaluated individually by a decision-maker "creditor". The success of a judgemental process depends on the

experience and the common sense of the credit analyst. As a result, judgemental techniques are associated with subjectivity, inconsistency and individual preferences motivating decisions; and judgemental methods have some strengths, such as taking account of qualitative characteristics and having a good track record in evaluating past credit by utilising the wealth of the credit analyst's past experience (see, for example, Al Amari, 2002; citing Chandler & Coffman, 1979).

Otherwise, in a credit scoring model, analysts usually use their historical experience with debtors to derive a quantitative model for the segregation of acceptable and unacceptable credit applications. Using a credit scoring system, a credit application is largely a self-operating process and consistently applied to all credit decisions. The scoring system is based on the addition or subtraction of a statistically extracted number of points relating to the applicant's score given to the predictor variables, such as time on a job or the number of credit sources used. As a result, it can be said that credit scoring enables advancers to assess the creditworthiness quickly. Moreover, credit scoring gives a chance to the advancers to improve customer services and the retention of sound customers. By using a statistically extracted cut-off credit score, an analyst can, of course, separate the acceptable from the unacceptable credit applicants. On the other hand, credit scoring has been criticized because of statistical problems with the data used to evolve the model, as well as assumptions of the particular statistical technique used to derive the point scores. 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; Bailey, 2004).

1.3. Benefits and criticism of credit scoring

Comparisons between credit scoring techniques and personal judgemental techniques have not been adequately and critically studied, and only a few researchers have compared these two credit tools, such as Chandler & Coffman (1979) who were pioneers in comparing credit scoring with personal credit judgement; and Crook (1996) who summarized many of their arguments and discussed the benefits of credit scoring relative to judgemental techniques for both the lender and borrower.

Benefits of credit scoring: credit scoring requires less information to make a decision, because credit scoring models have been estimated to include only those variables, which are statistically and/or significantly correlated with repayment performance; whereas judgemental decisions, *prima facie*, have no statistical significance and thus no variable reduction methods are available (Crook, 1996). Credit scoring models attempt to correct the bias that would result from considering the repayment histories of only accepted applications and not all applications. They do this by assuming how rejected applications would have performed if they had been accepted. Judgemental methods are usually based on only the characteristics of those who were accepted, and who subsequently defaulted (Crook, 1996). Credit scoring models consider the characteristics of good as well as bad payers, while, judgemental methods are generally biased towards awareness of bad payers only. Credit

scoring models are built on much larger samples than a loan analyst can remember. Credit scoring models can be seen to include explicitly only legally acceptable variables whereas it is not so easy to ensure that such variables are ignored by a loan analyst. Credit scoring models demonstrate the correlation between the variables included and repayment behaviour, whereas this correlation cannot be demonstrated in the case of judgemental methods because many of the characteristics which a loan analyst may use are not impartially measured. A credit scoring model includes a large number of a customer's characteristics simultaneously, including their interactions, while a loan analyst's mind cannot arguably do this, for the task is too challenging and complex. An additional essential benefit of credit scoring is that the same data can be analyzed easily and clearly by different credit analysts or statisticians and give the same weights. This is highly unlikely to be so in the case of judgemental methods (Chandler & Coffman, 1979; Crook, 1996).

Some other privileges of credit scoring has been summarised by Al Amari (2002), citing Chandler & Coffman (1979) as follows: more efficient processing time, and subsequent support for the decision-making process; minimization of credit process costs and effort; fewer errors made; provision of estimations to be compared in post audits; inclusion of variables supported through objective analysis to assess the credit risk; modelling based on real data; interrelation between variables are considered; fewer customer-information needs for credit decisions; the cut off score (more details of cut-off scores are given in the next sub-section) can be changed according to environmental factors affecting the banking sector; acceptance of only authorized factors considered by well known institutions, such as ECOA in the United States and Consumer Credit Act in the United Kingdom; and the choice of information supported as being related to customer credit risk.

Criticisms of credit scoring: credit scores use any characteristic of a customer in spite of whether a clear link with a likely repayment can be justified. Also, sometimes economic factors are not included. In addition, using credit scoring models, sometimes customers may have the characteristics, which make them more similar to bad than good payers, but may have these entirely by chance (a misclassification problem). Statistically a credit scoring model is "incomplete", for it leaves out some variables, which taken with the others, might predict that the customer will repay. But unless a credit scoring model has every possible variable in it, normally it will misclassify some people. Another criticism of credit scoring models is the possibility of indirect discrimination (Crook, 1996). Furthermore, credit scoring models: are not standardized and differ from one market to another; are expensive to buy and subsequently to train credit analysts; and sometimes a credit scoring system may 'reject (a) creditworthy applicant because he/she changes address or job' (Al Amari, 2002, p. 69; citing Chandler & Coffman, 1979).

Looking at how such credit scoring works, a customer for credit is evaluated in a credit scoring system by simply summing the points received on the various application features to have a total score. This score may be treated in a number of ways based on the system design. In a single cut-off method, the customer's total score is compared to a particular

single cut-off point score. If this score exceeds the cut-off, credit is awarded; otherwise the customer is denied. Advanced credit scoring systems are based on a two-stage process. For instance, the customer's total score may be compared to two cut-off points. If the score exceeds the higher cut-off, credit is automatically granted, while if it falls below the lower cut-off, credit is automatically rejected. If the score is between the two cut-off points, the analyst re-evaluates the customer based on the actual requirements, or alternatively "credit history information is obtained, scored, and the points added to the total score obtained (from information on the application form)". As a result of this approach, "if this new score is above a new higher cut-off, credit is (granted); if not, credit is denied" (Capon, 1982, p. 83).

Furthermore, credit scoring models are only as good as the original specification, and a further limitation is that the data are historical. Either the variables or weights, or both, are assumed to be constant over time, which makes the model less accurate, *unless it is frequently updated*. This problem can be reduced or even minimised if banks keep records of their type I and type II errors, and apply a new or up-dated model to address any necessary changes. A serious problem is that the model imposes a dichotomous outcome: either the borrower defaults or not. Indeed, a range of possible outcomes exists, from a delay in interest payments to non-payment of interest, to outright default on principal and interest. Often the borrower declares a problem with payments, and the loan terms can be renegotiated. These different outcomes can be included, but only two at a time (Heffernan, 2005).

2. Review of literature

2.1. How credit scoring has developed in importance

It is believed that credit scoring, regardless of all the criticisms, can seriously help to answer some key questions. However, Al Amari (2002, p. 41) has argued that while a lot of credit scoring models have been used in the field, these key questions have not been yet answered conclusively: What is the optimal method to evaluate customers? What variables should a credit analyst include to assess their applications? What kind of information is needed to improve and facilitate the decision-making process? What is the best measure to predict the loan quality (whether a customer will default or not)? To what extent can a customer be classified as good or bad?

In addition to Al Amari's questions the following can usefully be added: What is the best statistical technique on the basis of the highest average correct classification rate or lowest misclassification cost or other evaluation criteria? Can alternative credit scoring models offer the credit decision-makers more efficient classification results than judgmental approaches? Does the predicted credit quality based on conventional techniques adequately compare with those based on more advanced approaches? Is it possible to identify the key factors using credit scoring that can strongly influence loan quality? The latter have clearly been neglected in the literature except for Abdou (2009a), who argues that sophisticated credit scoring techniques can fully address these additional questions.

The role of effective management of different financial and credit risks is especially important for bankers, who have come to realise that banking operations affect and are affected by economic and social environmental risks that they face, and that consequently the banks might have an important role to play in helping to raise banking environmental requirements. Although the environment presents significant risks to banks, in particular environmental credit risk, it also perhaps presents profitable opportunities (Thompson, 1998; Casu, et al., 2006). The management of risk plays an important role in the banking sector worldwide. One of the key components of risk management is that associated with the personal credit decision. Indeed this is one of the most critical banking decisions, requiring a distinction between customers with good and bad credit. The behaviour of former and current customers can provide a useful historical data-set, which can be crucial in predicting new applicants' behaviour.

With the fast growth of the credit industry all over the world and portfolio management of huge loans, credit scoring is regarded as one of the most important techniques in banks, and has become a very critical tool during recent decades. Credit scoring models are widely used by financial institutions, especially banks, to assign credit to good applicants and to differentiate between good and bad credit. Using credit scoring can reduce the cost of the credit process and the expected risk associated with a bad loan, enhancing the credit decision, and saving time and effort (Lee et al, 2002; Ong et al, 2005). Decision-making involving 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 the past and present experiences of credit analysts whose evaluation of clients includes their ability to repay credit, guarantees and client's character (Sarlija, et al., 2004). Due to the rapid increase in fund-size invested through credit granted, and the need for quantifying credit risk, financial institutions including banks have started to apply credit scoring models.

A credit scoring system should be able to classify customers as good credit those who are expected to repay on time and as bad credit those who are expected to fail. Credit scoring, which helps to classify groups of customers correctly, can also assist banks in increasing sales of additional products. One of the main goals of credit scoring in financial credit institutions and banks is to help the development of the credit management process and to provide credit analysts and decision-makers with an efficient and effective credit tool to help to determine strengths, weaknesses opportunities and threats (SWOT); and to help to evaluate credit more precisely. A major problem for banks is how to determine the bad credit, because bad credit may cause serious problems in the future. This leads to loss in bank capital, lower bank revenues and subsequently increases bank losses, which can lead to insolvency or bankruptcy.

In developed countries, credit scoring is well established and the number of applications is increasing, because of excellent facilities and vast information being widely available, whilst in less developed or developing countries, less information and facilities are available. Advanced technologies, such as those used with credit scoring have helped credit analysts in

different financial institutions to evaluate and subsequently assess the vast number of credit applications. West (2000, p. 1132) has stated that credit scoring is widely used by the “financial industry”, mainly to improve the credit collection process and analysis, including a reduction in credit analysts’ cost; faster credit decision-making; and monitoring of existing customers. Also, around 97% of banks are using credit scoring for credit card applications, and around 82% of banks (and it was not clear from the original source whether the author was referring to US banks only) are using credit scoring to decide correctly who should be approved for credit card applications. Furthermore, credit institutions and especially mortgage organizations are developing new credit scoring models to support credit decisions to avoid large losses. These losses were considerable. For example, West (2000:1132) reported that ‘in 1991 \$1 billion of Chemical Bank’s \$6.7 billion in real estate loans were delinquent’.

Gathering information is a critical issue in building a credit scoring model. In general, through loan application forms, customer bank account(s), related sector(s), customer credit history, other financial institutions and banks, market sector analysis and through government institutions, banks may gain competitive advantages by building a robust credit scoring model(s). By collecting and isolating all relevant information, credit analysts or “banks” should be able to decide whether a particular variable should be included in the final model or not, and additionally whether a variable fits the real field requirements.

2.2. Credit scoring applications

As discussed earlier, the concept of credit scoring and related concepts of behavioural and profit scoring (see for example, Sarlija et al., 2009; Banasik and Crook, 2010; Fritz & Hosemann, 2000), are not that old, compared with credit and other business concepts. Nevertheless, applications of credit scoring have been widely used in different fields, including a comparison between different statistical techniques used in prediction purposes and classification problems. These applications can be classified into accounting and finance (Landajo et al, 2007; Pendharkar, 2005; Baestaens, 1999; Altman et al, 1994; Sinha and Richardson, 1996; Duliba, 1991; Long, 1973); marketing (Chiang et al, 2006; Thieme et al, 2000; Kumar et al, 1995; Dasgupta et al, 1994); engineering and manufacturing (Dvir et al, 2006; Feng & Wang, 2002; Smith & Mason, 1997); health and medicine (Behrman et al, 2007; Nguyen et al, 2002; Warner & Misra, 1996); and general applications (Nikolopoulos et al, 2007; Usha, 2005; Walczak & Sincich, 1999; Hardgrave et al, 1994), as noted by Paliwal & Kumar (2009).

In the area of accounting and finance, credit scoring applications have been used for different purposes, particularly with the rapid growth in this area. The number of applications has increased during the last couple of decades, such as bankruptcy prediction (Tsai & Wu, 2008; Etemadi et al, 2009; Min & Lee, 2008; Nanni & Lumini, 2009; Atiya, 2001) and bankruptcy classification (Min & Jeong, 2009; Lensberg et al, 2006; MaKee & Lensberg, 2002; Ignizio & Soltys, 1996); scoring applications (Crook et al, 2007; Huang et al, 2007; Huang et al, 2006), classification problems (Ben-David & Frank, 2009; Laha, 2007; Trinkle &

Baldwin, 2007; Lensberg et al, 2006; Beynon, 2005; Ong et al, 2005; Zhang & Bhattacharyya, 2004; Glen, 2001); financial distress (Hu, 2008; Mukkamala, et al., 2008; Hu & Ansell, 2007; Van Gestel et al, 2006) and financial decisions and financial returns (Yu et al, 2009; West et al, 2005; Xia et al, 2000).

Credit scoring applications in banking sectors have expanded during the last couple of decades (Banasik and Crook, 2010; Chen et al, 2009; Thanh Dinh & Kleimeier, 2007; Hand et al, 2005; Emel et al, 2003; Foglia et al, 1998), especially due to the large number of credit applications for different bank products, providing a wide range of new product channels which can be used by these banks. These applications and others have included different bank products, such as consumer loans, which are one of the most important and essential applications widely used in the field (Sustersic et al, 2009; Lee & Chen, 2005; Hsieh, 2004; Kim & Sohn, 2004; Malhotra & Malhotra, 2003; Orgler, 1971); credit card scoring applications, which are one of the earliest applications in the field (Quah & Sriganesh, 2008; Lee et al, 2002; Banasik et al, 2001; Greene, 1998); small businesses, which are another important bank product nowadays (DeYoung et al, 2008; Carter & McNulty, 2005; Bensic et al, 2005; Frame et al, 2004; Frame et al, 2001; Tsaih et al, 2004; Stefanowski & Wilk, 2001; Kumar & Motwani, 1999; Falbo, 1991); and other bank products, such as mortgages, which have recently been massively used by different banks as well (Haughwout et al, 2008; Somers & Whittaker, 2007; Heuson et al, 2001; Cameron & Trivedi, 1996).

Consumer credit has become a huge industry, and the number of applications has increased during the last couple of decades. In the late 1980s, in the United States, over \$700 billion was the total amount of outstanding consumer credit. Also Eastern European countries and China have started to see the usefulness of consumer credit, because a key task of consumer credit is to make credit extensively available and to make it profitable as well (Lewis, 1992, pp. 1-2). The evaluation of new consumer loans is one of the most important applications of credit scoring models and has attracted attention in the last few decades (Crook et al, 2007; Baesens et al, 2005; Hsieh, 2004; Sarlija et al, 2004; Malhotra, & Malhotra, 2003; Steenackers & Goovaerts, 1989). Some other researchers have focused on existing consumer loans rather than new loan applications (Kim & Sohn, 2004; Orgler, 1971). Feelders (2000) proposed a reject inference method based on mixture modelling, which led to significant improvement of the classification rule. Also, Banasik and Crook (2010) have analysed the procedure of augmentation to deal with reject inference in the context of high rejection rates, and have concluded that there are no positive benefits.

For other scoring applications, such as corporate credit scoring models (see for example Altman, 2005; Paleologo, et al. 2010), the nature and the requirements of the scoring system might be different and need to include a number of stages, such as those suggested by Altman & Haldeman (1995). These steps include: applying primary client-data to credit scoring models; testing a credit scoring model and using a supplemental system. For the second step, the model requires “tests covering the following issues: definition of risk, model development, test of time, stability, public versus private company data, probability of failure,

credibility, model support and pilot testing” (Altman & Haldeman, 1995, p. 13). Furthermore, the third step can be important for the model: “smoothing out the wave, firm- capital market approach” (i.e. using systematic beta risk) and “firm-econometric approach” (Altman & Haldeman, 1995, pp. 19-22). For further explanation of these stages, the reader is referred to Altman & Haldeman, (1995).

2.3. Key determinants of credit scoring

The objective of credit scoring models is to assign loan customers to either good credit or bad credit (Lee et al, 2002), or predict the bad creditors (Lim & Sohn, 2007). Therefore, scoring problems are related to classification analysis (Anderson, 2003; Lee et al, 2002; Hand, 1981). Probably the earliest use of statistical scoring to distinguish between “good” and “bad” applicants was by Durand (1941), who analysed data from financial services, such as commercial and industrial banks, and finance and personal finance companies. With the wide use of credit scoring in the United States and United Kingdom, and following the Equal Credit Opportunity Act (ECOA) in the United States, credit scoring models become well-known and the credit scoring literature has expanded. Credit scoring was primarily dedicated to assessing individuals who were granted loans, both existing and new customers. Credit analysts, based on pre-determined scores, reviewed customers’ credit history and creditworthiness to minimize the probability of delinquency and default (Al Amari, 2002).

The categorisation of good and bad credit is of fundamental importance, and is indeed the objective of a credit scoring model (Lim & Sohn, 2007; Lee et al, 2002). The need of an appropriate classification technique is thus evident. But what determines the categorisation of a new applicant? From the review of literature, characteristics such as gender, age, marital status, dependants, having a telephone, educational level, occupation, time at present address and having a credit card are widely used in building scoring models (Sustersic et al, 2009; Hand et al. 2005; Lee and Chen 2005; Sarlija et al., 2004; Banasik et al. 2003; Chen & Huang, 2003; Lee et al., 2002; Orgler 1971; Steenackers and Goovarts 1989). Time at present job, loan amount, loan duration, house owner, monthly income, bank accounts, having a car, mortgage, purpose of loan, guarantees and others have been also used in building the scoring models (Ong et al. 2005; Lee and Chen, 2005; Greene 1998; Sarlija et al., 2004; Orgler 1971; Steenackers and Goovarts 1989). In some cases the list of variables has been extended to include spouse personal information, such as age, salary, bank account and others (Orgler 1971). Of course, more variables are less frequently used in building scoring models, such as television area code, weeks since the last county court judgement, worst account status, time in employments, time with bank and others (Bellotti and Crook, 2009; Banasik and Crook, 2007; Andreeva, 2006; Banasik et al. 2003).

Insights can be gained from parallel research, pertaining to small business and corporate loans, by identifying other variables, such as main activity of the business, age of business, business location, credit amount, and different financial ratios, for example, profitability, liquidity, bank loans and leverage have been used in scoring applications (Emel et al. 2003;

Bensic et al, 2005; Zekic-Susac et al. 2004; Min and Lee, 2008; Min and Jeong, 2009; Lensberg et al. 2006; Cramer, 2004; Liang 2003).

In some cases the final selection of the characteristics was based on the statistical analysis used, i.e. stepwise logistic regression, regression or neural network (Lee and Chen, 2005; Nakamura, 2005; Kay & Titterington, 1999; Lenard, et al., 1995; Steenackers and Goovarts 1989; Orgler 1971). However, to the best of our knowledge, none of the research reviewed in this paper has clearly established a theoretical reason why such variables have been chosen. In addition, in most cases, authors have stated that a particular set of data was provided by a particular institution. Therefore, the selection of the variables used in building scoring models depends on the data providers and the data availability as stated by those authors. It is the view in this paper that such variables are implicitly deemed influential.

Classification models for credit scoring are used to categorize new applicants as either accepted or rejected with respect to these characteristics. These need to be contextualised to the particular environment, as new variables are appropriately included (see, for example, the inclusion of corporate guarantees and loans from other banks within the Egyptian environment in the investigation by Abdou and Pointon, 2009). The classification techniques themselves can also be categorised into conventional methods and advanced statistical techniques. The former include, for example, weight of evidence, multiple linear regression, discriminant analysis, probit analysis and logistic regression. The latter comprise various approaches and methods, such as, fuzzy algorithms, genetic algorithms, expert systems, and neural networks (Hand & Henley, 1997). On the one hand, the use of only two groups of customer credit, either “good” or “bad” is still one of the most important approaches to credit scoring applications (Kim & Sohn, 2004; Lee et al, 2002; Banasik et al, 2001; Boyes et al, 1989; Orgler, 1971). On the other hand, the use of three groups of consumer credit may become one of the approaches for classification purposes in credit scoring models. Some have used “good” or “bad” or “refused” (Steenackers & Goovaerts, 1989), whilst others have used “good” or “poor” or “bad” (Sarlija et al, 2004). Lim & Sohn (2007) argue that the way existing models are used is quite worrying, especially at the time when the middle of the repayment term occurs, when it is important to be able to re-evaluate the creditability of borrowers with high default risks for the remaining term (see for example, Baestaens, 1999).

It is believed that there is no optimal number of variables that should be used in building scoring models. The selection of the variables varies from study to study based on the nature of the data, and also on what cultural or economic variables may affect the quality of the model and be appropriate to a particular market, whose variables differ from one country to another. In finance applications, a rank from only three variables (Pendharkar, 2005; Fletcher & Goss, 1993) to around twenty variables (Jo et al, 1997; Desai et al, 1996; Tam & Kiang, 1992) has been used in building the scoring models. Meanwhile, others have used more variables in their analysis. For example, Salchenberger et al. (1992) applied twenty-nine variables, and forty-one variables have been used by Leshno & Spector (1996). In other fields' applications, such as health & medicine and engineering & manufacturing, the number

of variables has been vastly increased to seventy-two variables, in a study by Delen et al. (2005), and to eighty-five variables, in a study by Dvir et al. (2006), respectively.

While most of the authors are agreed about the importance of credit scoring methodology and the utmost necessity of developing a system “model” with a strong predictive ability, there has been disagreement about what is the most appropriate cut-off score in evaluating customer credit. The determination of the cut-off point(s) is central to the usefulness and value of credit scoring models. Depending on a pre-determined score point, a new customer can be classified as accepted or rejected, but to measure the score for this customer the cut-off point should be based on some technical factors. Different statistical techniques, such as discriminant analysis, logistic regression, neural networks and other statistical models have been widely used to evaluate consumer credit (Al Amari, 2002). Overall, there is no optimal cut-off score point. It varies from one environment to another and from one bank to another inside the same country. It is mainly based on the credit decision-makers’ attitudes to risk; in some cases a bank might be able to accept more clients, in this case the cut-off score point is expected to be lower than the one currently used, and vice versa.

Moreover, the determination of the sample size is another issue that has been discussed. It is believed that the larger the sample size, the better the scoring model’s accuracy. These determinations mainly depend on the data availability, the nature of the market and to what extent this particular data-set will represent the community. In the field of *finance*, some studies have utilized a very small number of observations in their analysis, around three or four dozen (Dutta et al, 1994; Fletcher & Goss, 1993), whilst others have employed a larger number of observations in their applications, indeed thousands of observations (Bellotti & Crook, 2009; Hsieh, 2004; Banasik et al, 2003). In some personal/consumer loan applications, a number of studies have applied their analysis based on data-sets of less than one thousand one hundred observations (Sustersic et al, 2009; Lee & Chen, 2005; Kim & Sohn, 2004; Malhotra & Malhotra, 2003). Furthermore, sample selection bias has been highlighted in other studies, particularly pertaining to the analysis of only applicants that have been accepted (Banasik & Crook, 2007; Banasik & Crook, 2005; Verstraeten & Van Den Poel, 2005; Banasik et al, 2003; Greene, 1998).

The categorization of the sample or “validation method” has been widely used in financial credit scoring applications. While some researchers have applied a simple validation technique by dividing the sample into training and testing/applied sub-samples (Landajo et al, 2007; Limsombunchai et al, 2005; Lee & Jung, 2000), other researchers have utilized a different validation technique by segregating the sample into training and validation and testing sub-samples (Spear & Leis, 1997; Salchenberger et al, 1992). Under the former validation technique, some studies have used a 50%-50% proportion between the two samples (Lenard et al, 1995; Yoon et al, 1993), whilst others have applied a 70%-30% or “2/3-1/3” proportion between different samples (Lee et al, 2002; Desai et al, 1996; Boritz & Kennedy, 1995; Dutta et al, 1994).

In addition, researchers have applied a different number of folds, i.e. 4-fold cross validation (CV) and 5-fold CV into their analysis (Lee et al, 2005; Zhang et al, 1999). Also, a 90%-10% validation procedure has been used in personal loans' (housing loans) financial applications (Lee & Chen, 2005). Other proportions, such as 80%-20%, have been used under other applications, such as marketing (Limsombunchai et al, 2005; Ainscough & Aronson, 1999), engineering and manufacturing (Heiat, 2002; Krishnaswamy & Krishnan, 2002). Paliwal & Kumar (2009) gave a summary of validation methods used in different credit scoring applications. In their review, it has been noted that thirty studies have split their data-set into training and testing data, five studies have split their data-set into three sub-samples (training, validation and testing), and seventeen studies have utilized n-fold cross validation. Other validation methods, such as bootstrapping, jack-knifing, and 'leave-one-out cross validation' were less frequently used compared with formerly cited methods.

In summary, it would seem from this discussion that there is no optimal credit scoring model procedure, including specific variables or number of variables, particular cut-off point, exact sample size and meticulous validation, which can be applied to different banks in different environments. This was also the conclusion reached by other authors, e.g. Al Amari (2002), who came to a similar conclusion that there is no best scoring model holding explicit variables that can be used in different markets; and Altman & Haldeman (1995, p. 22), in which they suggested nine variables for a logical scoring model, but finally they emphasised that, over time, a successful credit scoring system should be "credible and accepted by both senior management and the field troops, including loan officers and credit analysts". Therefore, a sound scoring model should reproduce the risk strategy and the credit culture of the institutions.

2.4. Credit scoring statistical techniques

A wide range of statistical techniques are used in building the scoring models. Most of these statistical, and some of these non-linear, models are applicable to build an efficient and effective credit scoring system that can be effectively used for predictive purposes. Techniques, such as weight of evidence measure, regression analysis, discriminant analysis, probit analysis, logistic regression, linear programming, Cox's proportional hazard model, support vector machines, decision trees, neural networks, *k*-nearest-neighbour, genetic algorithms and genetic programming, are all widely used techniques in building credit scoring models by credit analysts, researchers, lenders and computer software developers and providers.

Advanced statistical methods vs. traditional statistical methods: advanced statistical techniques, such neural networks and genetic programming provide an alternative to conventional statistical techniques, such as discriminant analysis, probit analysis and logistic regression. The point of using sophisticated techniques, such as neural nets, is their capability of modelling extremely complex functions, and, of course, this stands in contrast to traditional linear techniques, such as, linear regression and linear discriminant analysis.

Probabilistic neural nets usually trains presented cases faster than multi-layer feed-forward nets, and classifies them in the same way or better than multi-layer feed-forward nets, even through multi-layer feed-forward nets have been shown to be excellent classifiers (Palisade, 2005; Irwin, et al., 1995). However, a range of sophisticated algorithms for neural nets training, making them an attractive alternative to the more conventional techniques, has become available (Masters, 1995; Palisade, 2005). Also, genetic programming is one of the most successful alternatives to traditional techniques recently used in the field. Genetic programming is utilized to automatically determine the sufficient discriminant functions and the applicable features simultaneously. Dissimilar neural networks may only suit large data-sets, but genetic programming can positively; perform well even with small data-sets (Nath et al, 1997). It is useful to discuss some of the previously mentioned credit scoring modelling techniques as follows:

Linear regression methods have become an essential component of any data analysis concerned with describing the relationship between a response variable and one or more independent variables. Linear regression has been used in credit scoring applications, as the two class problem can be represented using a dummy variable. Using a Poisson regression model instead could be used to accommodate cases where the customer makes varying degrees of partial repayments. As such the proportionate repayments could be re-expressed as Poisson 'counts'. Factors, such as customers' historical payments, guarantees, default rates in a timely manner, can be analysed by credit analysts, with linear regression to set up a score for each factor, and then to compare it with the bank's cut-off score. If a new customer's score passes the bank's score, the credit will be granted. Orgler (1970) used regression analysis for commercial loans; this model was limited to the evaluation of existing loans and could be used for loan review and examination purposes. Later on, Orgler (1971) used a regression approach for evaluating outstanding consumer loans. He came to the conclusion that information not included on the application form had greater predictive ability than information included on the original application form, in assessing future loan quality. The use of regression analysis extended such applications to include further aspects (see, Lucas, 1992; Henley, 1995; Hand & Henley, 1997; Hand & Jacka, 1998).

Discriminant analysis is a simple parametric statistical technique, developed to discriminate between two groups. Many researchers have agreed that the discriminant approach is still one of the most broadly established techniques to classify customers as good credit or bad credit. This technique has long been applied in the credit scoring applications under different fields. Therefore, credit scoring model based on a discriminant approach is basically used for statistical analysis to classify groups' variables into two or more categories. Discriminant analysis was first proposed by Fisher (1936) as a discrimination and classification technique. 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. Indeed, discriminant analysis is a valid technique used in building credit scoring models (Abdou & Pointon, 2009; Sarlija et al, 2004; Caouette et al, 1998; Hand et al, 1998; Hand & Henley, 1997; Desai et al, 1996).

Several authors have expressed pointed criticism of using discriminant analysis in credit scoring. Eisenbeis (1978) noted a number of the statistical problems in applying discriminant analysis based on his earlier work in 1977. Problems, such as using linear functions instead of quadratic functions, groups' definition, prior probabilities inappropriateness, classification error prediction and others, should be considered when applying discriminant analysis. Regardless of these problems, discriminant analysis is still one the most commonly used techniques used in credit scoring (Abdou et al. 2009; Greene, 1998).

Probit analysis is another conventional technique used in credit scoring applications for many years. Grablowsky & Talley (1981, p. 254) noted that probit analysis was first pioneered for the analysis of "toxicology problems" by Finney (1952), who used it to "determine the relationship between the probability that an insect will be killed and the strength of the dose of poison administrated". However, early in the 1930s the term "Probit" was developed and stood for probability unit (Pindyck & Rubinfeld, 1997; Maddala, 2001). Probit analysis is a technique that finds coefficient values, such that this is the probability of a unit value of a dichotomous coefficient. Under a probit model, a linear combination of the independent variables is transformed into its cumulative probability value from a normal distribution.

Grablowsky & Talley (1981, p. 260) stated that, under probit analysis, normal distributions of the "threshold values" are assumed, while multivariate normal distributions and equal variances are assumed under discriminant analysis; and using a likelihood ratio test, estimates of coefficients under a probit function can be tested individually for significance because of their 'uniqueness'. But, this is not the case for discriminant coefficients, which cannot be individually tested, whilst this is possible in a regression as well as under a probit function, but the latter is much more difficult than that for a linear, logistic or Poisson regression model. Finally, they note that multicollinearity can cause, under probit analysis, incorrect signs for coefficients, although the probability values from the likelihood ratio tests are not affected. Otherwise, this problem is not an issue under discriminant analysis.

TABLE 1 HERE

Table 1 reveals classification results of different scoring models investigated by Guillen & Artis (1992). The first column shows the total correct classification, the second column is the correct classification of good, the third column is the correct classification of bad, and the fourth column is the percentage of bad accepted into the good group. It can be observed from Table 1 that the probit model has the highest correct total classification rate of 71.9%. Yet, it has the worst rate for classifying bad cases accepted in a good group (i.e. type II error), which are serious misclassifications in practice because of the default implications. By contrast the

linear regression model has the lowest bad cases accepted in a good group even though its total correct classification rate is the worst amongst all models. It would be more meaningful to calculate both the type I and type II errors, applying a cost function to each on account of the different associated opportunity costs and produce an overall misclassification score, choosing the optimal model as the one with the lowest misclassification cost (see West, 2000; Abdou and Pointon, 2009; Abdou et al. 2009).

One of the other techniques used in credit scoring applications, is the *weight of evidence measure*. While a few numbers of studies have investigated the use of the weight of evidence measure in the field, results were comparable with those from other techniques (Abdou, 2009b; Banasik et al, 2003; Bailey, 2001; Siddiqi, 2006). The use of probit analysis has also been investigated as well, and compared with other statistical scoring models (Abdou, 2009c; Guillen & Artis, 1992; Banasik et al, 2003; Greene, 1998); also classification results were very close to other techniques (Greene, 1998), and better than techniques, such as discriminant analysis, linear regression and the Poisson model (Guillen & Artis, 1992). Furthermore, probit analysis is used as a successful alternative to logistic regression.

Logistic regression, like discriminant analysis, is also one of the most widely used statistical techniques in the field. What distinguishes a logistic regression model from a linear regression model is that the outcome variable in logistic regression is dichotomous (a 0/1 outcome). This difference between logistic and linear regression is reflected both in the choice of a parametric model and in the assumptions. Once this difference is accounted for, the methods employed in an analysis using logistic regression follow the same general principles used in linear regression (Hosmer & Lemeshow, 1989). The simple logistic regression model can easily be extended to two or more independent variables. Of course, the more variables, the harder it is to get multiple observations at all levels of all variables. Therefore, most logistic regressions with more than one independent variable are done using the maximum likelihood method (Freund & William, 1998). On theoretical grounds it might be supposed that logistic regression is a more proper statistical instrument than linear regression, given that the two classes “good” credit and “bad” credit have been described (Hand & Henley, 1997). Logistic regression has been extensively used in credit scoring applications (see for example: Abdou, et al., 2008; Crook et al, 2007; Baesens et al, 2003; Lee & Jung, 2000; Desai et al, 1996; Lenard et al, 1995).

In building the scoring models, statistical techniques such as discriminant analysis, regression analysis, probit analysis and logistic regression, have been evaluated (Sarlija et al, 2004; Banasik et al, 2001; Greene, 1998; Leonard, 1992; Steenackers & Goovaerts, 1989; Boyes et al, 1989; Orgler, 1971). Other methods are: mathematical programming, non-parametric smoothing methods, Markov chain models, expert systems, neural networks, genetic algorithms and others (Hand & Henley, 1997). Also, case studies have been the subject of investigation in the credit scoring literature (see, for example: Lee & Chen, 2005; Lee et al, 2002; Banasik et al, 2001; Leonard, 1995; Myers & Forgy, 1963).

Decision trees are another classification techniques used in developing credit scoring models, also known as recursive partitioning (Hand & Henley, 1997) or Classification and Regression Trees (CART). Probably one of the first uses of a CART model was pioneered by Breiman et al. (1984). However, Rosenberg & Gleit (1994) stated that the first model based on a decision tree was initiated by Raiffa & Schlaifer (1961) at the Harvard Business School, and also stated that later on a credit scoring model derived from decision trees was developed by David Sparks in 1972 at the University of Richmond. A classification tree is a non-parametric method to analyse dependent and/or categorical variables as a function of continuous explanatory variables (Breiman et al. 1984; Arminger et al, 1997). In a classification tree, a dichotomous tree is built by splitting the records at each node based on a function of a single input. The system considers all possible splits to find the best one, and the winning sub-tree is selected based on its overall error rate or lowest cost of misclassification (Zekic-Susac et al, 2004). A comparison of discriminant analysis and recursive partitioning was investigated by Boyle et al. (1992). Other applications of decision trees in credit scoring were described by Baesens et al. (2003), Stefanowski & Wilk (2001), Thomas (2000), Fritz & Hosemann (2000), Hand & Jacka (1998), Henley & Hand (1996), and Coffman (1986). Also, Paleologo et al. (2010) evaluate credit requests from corporate clients, address the issue of unbalanced data sets, and use a subagging procedure within their decision tree paradigm which utilizes extreme values for missing data.

TABLE 2 HERE

Table 2 summarises a comparison between decision trees and other techniques, such as logistic regression and *K*-nearest neighbour (*K*-NN), in terms of average bad risk rate, by Henley & Hand (1996). The bad risk rates were clearly similar for the different scoring techniques. It is also clear that this study had a much higher proportion of bad rates than other studies.

More sophisticated models, also known as artificial intelligence include, for example, expert systems, neural networks and genetic programming (see for example, Sustersic et al, 2009) are discussed below.

Expert systems are one of the new technologies recently applied into credit scoring applications, which depend on human experts' knowledge, interpretation and way of thinking to solve complex problems (Rosenberg & Gleit 1994). Research on expert systems, in this context, is so limited and unfortunately does not provide much detail. Hand & Henley (1997) noted that one of the expert systems' privileges is the ability to explain outcomes and, of course, this can provide reasons for denying a credit applicant. Rosenberg & Gleit (1994, p. 601) briefly discussed what Nelson & Illingworth (1990) stated about the main three components of such an expert system, which is relying on knowledge, which includes "facts and rules", whose combination requires a conclusion, by an engine, and "an interface" to

enable users to understand and, therefore, explain decisions and recommendations, and then it updates this information.

Recently, some other applications using expert systems have been published. They include the work by Ben-David & Frank (2009), who made a comparison between machine learning models and a credit scoring expert system, whose results revealed that while some of the machine learning models' accuracies are better than those expert system model, most of them are not; Kumra et al. (2006) applied an expert system approach to a commercial loan, and found that the expert system can introduce many characteristics of the "underwriting process" that different approaches do not (for other earlier applications, see Lovie, 1987; Leonard, 1993).

Neural networks are mathematical techniques motivated by the operations of the human brain as influential in problem solving techniques. Gately (1996, p. 147) defined neural networks as "an artificial intelligence problem solving computer program that learns through a training process of trial and error". Therefore, neural networks' building requires a training process, and the linear or non-linear variables in the training procedure help distinguish variables for a better decision-making outcome. In the credit scoring area, neural networks can be distinguished from other statistical techniques. Al Amari (2002, p. 63) gave an example to differentiate between regression models and neural networks models. In his discussion, he stated that to build an applicant score using regression models, the "inverse matrix" should be used, whilst in neural networks the "applicants' profile" is used to perceive those applicants' relative scores. Also, using neural networks, if the outcomes are unacceptable, the estimated scores will be changed by the nets until they become acceptable or until having each applicant's optimal score.

Recently neural nets have emerged as a practical technology, with successful applications in many fields in financial institutions in general, and banks in particular. Applications, such as credit card fraud, bankruptcy prediction, bank failure prediction, mortgage application, option pricing and others were suggested by Gately (1996) as financial areas where neural networks can be successfully used. They address many problems, such as pattern recognition, and make use of feed-forward nets' architecture, such as the multi-layer feed-forward nets and probabilistic neural networks, representing the majority of these applications (Bishop, 1995; Masters, 1995). A few credit scoring models using probabilistic neural nets have been investigated (Masters, 1995; Zekic-Susac et al, 2004).

TABLE 3 HERE

Correspondingly, of course, many scoring models applying multi-layer feed-forward nets have been used (Dimla & Lister, 2000; West, 2000; Reed & Marks, 1999; Desai et al, 1996; Bishop, 1995; Trippi & Turban, 1993). The neural network models have the highest ACC rates in these studies when compared with discriminant analysis and logistic regression, or other techniques, although results are often very close. Table 3 summarises a comparison between

two types of neural networks and two conventional techniques, in terms of ACC rates by Abdou & Pointon (2009). The ACC rates were clearly better under neural network models compared with conventional models under different sub-samples.

Hybrid models, as well as neural networks and advanced statistical techniques have been used in building scoring models (Trinkle & Baldwin, 2007; Blochlinger & Leippold, 2006; Seow & Thomas, 2006; Lee & Chen 2005; Yim & Mitchell, 2005; Kim & Sohn 2004; Lee et al, 2002; Stefanowski & Wilk, 2001). Meanwhile, comparisons between traditional and advanced statistical techniques have been investigated too (Abdou & Pointon, 2009; Abdou et al. 2009; Lee & Chen 2005; Ong et al, 2005; Zekic-Susac et al, 2004; Malhotra & Malhotra, 2003; Lee et al, 2002; Fritz & Hosemann, 2000). Comparisons have also been extended to include feed-forward nets and back-propagation nets (Malhotra & Malhotra, 2003; Armingier et al, 1997). Statistical association measures showed that the neural network models are better representations of data than logistic regression and CART (Zekic-Susac et al, 2004), while discriminant analysis, in general, has a better classification ability but worse prediction ability, whereas logistic regression has a relatively better prediction capability (Liang, 2003). Generally, the neural network models have the highest average correct classification rate when compared with other traditional techniques, such as discriminant analysis and logistic regression, taking into account the fact that results were very close (see, for example, Abdou, et al., 2008; Crook et al, 2007; Zekic-Susac et al, 2004; Haykin, 1994).

TABLE 4 HERE

West (2000, p. 1150) has developed five different neural networks' architectures, using German and Australian credit scoring data-sets. Based on West's credit scoring error analysis' results, it has been suggested that both "the mixture-of-experts (MOE) and radial basis function (RBF) neural networks should be considered for scoring applications", whilst multilayer perceptron (MLP) may not be the utmost precise neural net model. Also, logistic regression is considered as the most accurate model between conventional models, as shown in Table 4.

Genetic programming is one of the most recent techniques that has been applied in the field of credit scoring. It began as a subset of genetic algorithmic techniques, and can be considered as an extension of genetic algorithms (Koza, 1992; Goldberg, 1989). *Genetic algorithms* transform a data-set according to fitness value, by applying genetic operations. Under genetic algorithms, the solution is in the form of a "string" (Kaza, 1992). In genetic programming a set of competing programs are randomly generated by processes of mutation and crossover, which mirror the Darwinian theory of evolution, and the resultant programs are evaluated against each other. Generally, genetic programming generates competing programs in the LISP (or similar) language as a solution output (Nunez-Letamendia, 2002; Koza, 1994). The use of genetic programming applications is a rapidly growing area (Chen & Huang, 2003; Teller & Veloso, 2000), and the number of applications has increased during

the last couple of decades, such as bankruptcy prediction (Etemadi et al. 2009; MaKee & Lensberg, 2002), scoring applications (Huang et al. 2007; Huang et al. 2006), classification problems (Lensberg et al. 2006; Ong et al. 2005; Zhang & Bhattacharyya, 2004) and financial returns (Xia et al. 2000).

TABLE 5 HERE

Table 5 sums up predictive classification results of two genetic programming models (best genetic programme, GP_p , and best genetic team, GP_t) and two conventional techniques (weight of evidence and probit analysis), investigated by Abdou (2009c). It is clear that for the testing sample the classification results for genetic models were better than those for the weight of evidence model, whilst the results were comparable with probit analysis. Nevertheless the extra small percentage point superiority of genetic programming may, for a large bank, be very valuable in terms of after-tax profit. For the overall sample, it is evident that genetic programming results were better than those for the conventional techniques (85.82% for GP_t which exceeds 81.93% for probit analysis).

Crook et al. (2007) summarize the predictive accuracy of different classifiers using credit scoring application data. Table 6 shows some of those studies' published results. It can be concluded from the results in Table 6 that there is no best credit scoring technique for all data-sets, it mainly depends on the details of the problem, the data structure and size, the variables used, the market for the application, and the cut-off point. Generally, the overall performance of advanced statistical techniques, such as neural nets and genetic programming, is better than other statistical techniques. Nevertheless, there is a role for conventional techniques, such as linear discriminant analysis and logistic regression in some studies. As noted by Crook et al. (2007), the figures in Table 6 can only be compared down a column, not between different studies. The reason is that these studies differ in how the cut-off was set, figures are not weighted according to the relative cost, and few studies have used statistical "inferential" tests to investigate if differences were significant.

TABLE 6 HERE

Most studies that have made a comparison between different techniques found that sophisticated statistical techniques such as neural networks, genetic programming and fuzzy algorithms are better than the traditional ones based on the average correct classification rate criterion. This sometimes depends on the original group that is used to compute the correct classification, depending on "bad" or "good and bad" together (Hoffmann et al. 2007; Blochlinger & Leippold, 2006; Desai et al. 1996). 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 the majority of cases not statistically different from other techniques (Baesens et al, 2003). It should be stressed that other statistical

techniques, such as support vector machines (see for example Deschaine & Francone, 2008), smoothing non-parametric methods, time varying models, mathematical programming, K-nearest neighbour, fuzzy rules, kernel learning method, Markov models and linear programming, have been discussed in the literature (see for example: Bellotti & Crook, 2009; Elliott & Filinkov, 2008; Crook et al., 2007; Hoffmann et al., 2007; Huang et al., 2007; Baesens et al., 2003; Yang, 2007; Hand & Henley, 1997).

2.5. Credit scoring performance evaluation criteria

Performance evaluation criteria, such as the confusion matrix or the Average Correct Classification (ACC) rate, the estimated misclassification cost, mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), the receiver operating characteristics (ROC) curve, GINI coefficient, and other criteria are all used in credit scoring applications under different fields. The following is a discussion of some of these performance criteria.

Confusion matrix (average correct classification rate criterion) is one of the most widely used criteria in the area of accounting and finance (for credit scoring applications) in particular, and other fields, such as marketing and health in general. The average correct classification rate measures the proportion of the correctly classified cases as good credit and as bad credit in a particular data-set. The average correct classification rate is a significant criterion in evaluating the classification capability of the proposed scoring models. The idea of correct classification rates comes from a matrix, which is occasionally called “a confusion matrix” (Zheng et al., 2004), otherwise called a classification matrix (Abdou, 2009c). A classification matrix presents the combinations of the number of actual and predicted observations in a data-set. In Yang et al. (2004) study, the confusion matrix was compared with another two criteria: Mahalanobis Distance and Kolmogorov-Smirnov Statistics with reference to ROC curve. In other studies this matrix has been compared with MSE and RMSE (Kumar et al. 1995; Fletcher & Goss, 1993). Commonly the mainstream of credit scoring applications either in accounting and finance or other fields have used the average correct classification rate as a performance evaluation measure (Paliwal & Kumar, 2009).

It is believed that the average correct classification rate is an important criterion to be used, especially for new applications of credit scoring, because it highlights the accuracy of the predictions. Yet, the ACC rate criterion does not accommodate differential costs to a bank, arising from different types of error. Specifically, it ignores different misclassification costs for the actual good predicted bad and the actual bad predicted good observations. In the real field it is believed that the cost associated with Type II errors is normally much higher than that associated with Type I errors (Baesens et al, 2003), as explained in the next section.

The estimated misclassification cost criterion simply measures the relative costs of accepting customer applications for loans that become bad versus rejecting loan applications that would be good. It is based on the confusion matrix; this criterion gives an evaluation of the effectiveness of the scoring models' performance, which can cause a serious problem to

the banks in the case of the absence of these estimations, especially with the actual bad predicted good observations. The estimated misclassification cost criterion, is a crucial criterion to evaluate the overall credit scoring effectiveness, and to find the minimum expected misclassification cost for the suggested scoring models.

A few credit scoring applications have used the estimated misclassification cost criterion in the field of finance (Abdou, 2009b; Abdou & Pointon, 2009; West, 2000; Lee & Chen, 2005) and in other fields (Hill & Remus, 1994). The reason, as noted by Lee & Chen (2005), is that the trustworthy or consistent estimates of the misclassification costs are a complicated and real challenging job to be provided, and, therefore, valid prediction might not be available. Lee & Chen (2005, p. 749) stated that “it is generally believed that the costs associated with (both) Type I error (good credit misclassified as bad credit) and Type II error (bad credit misclassified as good credit) are significantly different” and “the misclassification costs associated with Type II errors are much higher than (the misclassification cost) associated with Type I errors”. 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, which has been used by Abdou (2009b) as well. The use of this relative cost ratio has been extended, *in terms of sensitivity analysis*, to higher cost ratios (i.e. 7:1, 10:1 etc) as noted by Abdou (2009c) and Abdou & Pointon (2009).

The Receiver Operating Characteristics (ROC) curve, or sometimes called “Lorentz diagram”, is a two-dimensional graph, which represents the proportion of bad cases classified as bad (called ‘sensitivity’ which is plotted on the vertical axis) versus the proportion of good cases classified as bad (called ‘1 – specificity’ which is plotted on the horizontal axis) at all cut-off score values. In fact, sensitivity is equal to 1 minus the Type II error rate, and specificity is equal to 1 minus the Type I error rate, as shown in Figure 1 (Crook et al. 2007; Yu et al. 2009; Yang et al. 2004; Baesens et al. 2003). The ROC curve illustrates the achieved overall performance with reference to all cut-off score points. The ROC curve illustrates the behaviour of classifiers with no regard to misclassification costs or different class distributions; therefore, it effectively separates classification performance from these features (Yang et al. 2004; Baesens et al. 2003; Thomas et al. 2002). The ROC curve identifies appropriate cut-off score points, whose scores can maximize the Kolmogorov-Smirnov statistic, but it visualizes the details from the Kolmogorov-Smirnov statistic if the ROC is illustrated (Blochlinger & Leippold, 2006; Hand & Jacka, 1998).

FIGURE 1 HERE

Blochlinger & Leippold, (2006, p. 853) stated that “The maximum distance between the ROC curve and the diagonal equals a constant times the Kolmogorov-Smirnov statistic, but only if the ROC is concave. If the ROC curve is not concave, there is no such general correspondence”. The ROC curve was originally used in psychology, health and medicine, and manufacturing, as a technique to measure the performance of the “signal recovery

techniques” and “diagnostic systems”. Recently the ROC curve has been widely used in medicine and health applications (Song et al. 2005; Ottenbacher et al. 2004; Shang et al. 2000). Other fields, such as an engineering application, have witnessed the use of the ROC curve (Yesilnacar & Topal, 2005). Also, the use of the ROC curve in finance and banking applications has been observed (Banasik & Crook, 2007; Blochliger & Leippold, 2006; Baesens et al. 2003).

It should be emphasised that there are other performance evaluation criteria, such as the GINI coefficient, which “gives one number that summarizes the performance of the scorecard over all cut-off scores” (Thomas et al. 2002, p. 116), MSE, RMSE, MAE, and Goodness of Fit test (calibration). Table 7 summarizes some of the performance evaluation criteria investigated by Paliwal & Kumar (2009). It is clear from their review article that the most frequent performance criterion is the confusion matrix; and 18 out of their 36 cited studies are accounting and finance applications, whilst the remainder are in other fields. In terms of error rates, 25 studies used either mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE) or mean error,, and only 7 used the ROC curve.,

TABLE 7 HERE

Thus, there is no study, to the best of our knowledge, which has identified the optimal evaluation criterion. The best, in our opinion, would be determined by an array of factors, inter alia: the methodology used in the analysis, the nature of the data, the market where these data are collected, and the availability of the technology facilitating the analysis of very large data-sets.

At the practical level the choice of technology will depend on specific circumstances, for example, on whether it is a matter of intentional fraud or a matter of financial failure. Each of these requires slightly different mechanisms for detection. For the former, technologies such as artificial intelligence techniques (neural networks, data mining, genetic algorithms, fuzzy systems etc.) are good in detecting variations in customers’ behaviors. For the latter, close monitoring (such as period analysis) of customers’ financial portfolio, and a systematic breakdown of customers’ assets and liabilities may be needed. Of course, the big challenge here is that customers’ financial fortunes may be subjected to sudden changes such as bad investments (e.g. investors of Lehman Brothers) or plunge in value of financial assets (e.g. U.S. housing woes).

In this era of shortening economic cycle, values of financial assets may swing wildly at times and this make credit scoring of customers very challenging as it is almost always done behind the curve? The issue here is: can we make credit scoring be able to catch up with the dynamism of rapidly changing customers’ profiles? Obviously this can only be done using a combination of approaches, as follows: (i) sharing of customers’ financial profiles between lenders via credit bureau. This may be subjected to restrictions on banking secrecy requirements (ii) dynamically track spending patterns of customers. With increasing retailers

adopting on-line real-time system connected to the banking networks, this is a promising direction in catching unexpected customers' behaviors – frequently link to deteriorating credit profiles (iii) collateralized credit for customers with detected weak financial profiles. However, this has to be done discreetly as it may jeopardize customer relationship. At the end of the day, the lenders will have to strike a balance between being cautious and business expansion. Prudence is a delicate balancing act.

3. Conclusion

In this paper, we have carried out a comprehensive review of 214 studies in credit scoring, various performance evaluation criteria and different statistical techniques, which are used particularly in finance and banking. It has been settled in the literature that using scoring in credit evaluation rules out personal judgement. Credit scoring systems are numerical systems, and the decision will be taken, depending on the applicant's total score, whilst in personal judgement this issue is neglected, the decision here depends on decision-makers' personal experience and other cultural issues, which vary from market to market. It should be emphasised that there is no ideal credit scoring modelling procedure, which would guide the user in the choice of specific variables, cut-off score, validation method and sample size. It is not entirely clear how those factors may have had an influence on the alleged superiority of one technique over another, with ramifications or predictive ability in different circumstances. Our review clearly points out the key role of statistical scoring techniques in their use as a critical tool for prediction and classification problems. This review of the literature leads to the conclusion that there is no overall best statistical technique/method used for building credit scoring models, and the best technique for all data sets does not exist yet. As Hand & Henley (1997, p.535) conclude: what is best depends on the details of the problem, the structure of the data, the features of the application, the extent to which it is possible to segregate the classes by using those features, and the classification's objective(s).

Furthermore, a comparison between different statistical approaches demonstrates that advanced/sophisticated techniques, such neural networks and genetic programming perform better than more conventional techniques, such as discriminant analysis and logistic regression, in terms of their higher predictive ability. However, the results of some studies revealed that the predictive capabilities of both approaches were sufficiently similar to make it difficult to distinguish between them. These statistical techniques help credit decision-makers to predict banks' current and/or new customers as either good credit or bad credit, based on their attributes and "credit" information, and these performance evaluation criteria have also helped them to choose the best model based on their aims and objectives, constrained by their currently used evaluation system, specific inputs and target outcomes. However, misclassification costing is not a test of predictive capabilities, but an evaluation of the implications for the bank's costs. Misclassification costs are particularly important especially for type II errors, which misclassify bad loans as good. In reality it is difficult for researchers, although it is easier for the bankers, to establish more accurate costs. By contrast, type I

errors refer only to the opportunity cost of lost interest, whereas for type II misclassifications the bank loses some or all of not only the interest but also the repayment of principal. More recently, ROC and GINI, which are more advanced than other performance evaluation criteria, have been used. Although the GINI gives a measure of performance as a single score, the ROC provides useful information of the relative propensity of the two main misclassifications at different cut-off points. The banks should take their own evaluations of differential misclassification costs and use the ROC information to choose a cut-off point which minimizes the total misclassification costs. Credit scoring techniques are an astonishingly useful tool, which should help banks control an array of risks. It can be concluded that credit scoring developments and applications continue to be hugely expanded in various fields particularly in finance and banking. Also, the use of hybrid methods, such as the hybrid neural discriminant techniques offers one promising avenue for better classification and predictive capabilities.

This paper addresses a number of directions for future research. Firstly, having reviewed such a vast amount of literature on credit scoring, it seems surprising to observe that the ranking of the importance of variables used in building the scoring models are almost totally neglected in published research papers on credit scoring. This has important ramifications for the policies of the banks and for the banking system as a whole. Future research might usefully be employed in investigating this further. One of the reasons why the banks may not publish their own list of important variables may be because of their market image or ethical implications of their policies. Secondly, researchers in one discipline tend to ignore research in other disciplines partly because of salience and time pressures. However, research into personal bankruptcies from a social science perspective may throw light upon credit scoring. Future research should address the identification of drivers of default from a behavioural perspective, and the reasons for, inter alia, trends in self-bankruptcy determination, house repossession, rising education costs, and healthcare cost issues. Thirdly, not only does technology have implications for new modelling procedures, but the changing technological environment affects consumer spending patterns and the types of loans that they may wish to acquire, and consequently the types of loans that may be subject to default. Fourthly, researchers need to be innovative in establishing potentially important variables, as social and economic conditions change, in their credit scoring modelling procedures. Fifthly, research should focus more upon the timing of default within the period of the loan, and distinguish also between slow payers, intermittent payers and defaulters. Sixthly, and finally, in future research there needs to be incorporated into the modelling procedures time series aspects, so that trends in variable impact can be predicted. This is especially important for loans of longer duration, whose default is likely to be associated with differing attributes from those of short loans in a rapidly changing economic and social environment.

Acknowledgement

The authors would like to thank the editor and anonymous referees for helpful comments, which have been useful in revising the manuscript. All remaining errors are the authors' sole responsibility.

References

- Abdou, H. 2009a. Credit scoring models for Egyptian banks: neural nets and genetic programming versus conventional techniques, Ph.D. Thesis, The University of Plymouth, UK.
- Abdou, H. 2009b. An evaluation of alternative scoring models in private banking. *Journal of Risk Finance* 10 (1): 38-53.
- Abdou, H. 2009c. Genetic programming for credit scoring: The case of Egyptian public sector banks. *Expert Systems with Applications* 36 (9): 11402-11417.
- Abdou, H., Pointon, J. 2009. Credit scoring and decision-making in Egyptian public sector banks. *International Journal of Managerial Finance* 5 (4): 391-406.
- Abdou, H., Pointon, J., El Masry, A. 2008. Neural nets versus conventional techniques in credit scoring in Egyptian banking. *Expert Systems with Applications* 35 (3): 1275-1292.
- Ainscough, T. L., Aronson, J. E. 1999. An empirical investigation and comparison of neural networks and regression for scanner data analysis. *Journal of Retailing and Consumer Services* 6 (4): 205–217.
- Al Amari, A. 2002. The credit evaluation process and the role of credit scoring: A case study of Qatar. Ph.D. Thesis, University College Dublin.
- Altman, E. I. 2005. An emerging market credit scoring system for corporate bonds. *Emerging Markets Review* 6 (4): 311-323.
- Altman, E. I., Haldeman, R. 1995. Corporate credit scoring models: Approaches and tests for successful implementation. *Journal of Commercial Lending* 77 (9): 10-22.
- Altman, E. I. 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance* XXIII (4): 589-609.
- Altman, E. I., Marco, G., Varetto, F. 1994. Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking and Finance* 18 (3): 505–529.
- Anderson, R. 2007. *The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation*. New York: Oxford University Press.
- Anderson, T. W. 2003. *An Introduction to Multivariate Statistical Analysis*. New York: Wiley-Interscience.
- Andreeva, G. 2006. European genetic scoring models using survival analysis. *Journal of the Operational Research Society* 57(10): 1180-1187.

- Arminger, G., Enache, D., Bonne, T. 1997. Analyzing Credit Risk Data: A Comparison of Logistic Discriminant, Classification Tree Analysis, and Feedforward Networks. *Computational Statistics* 12 (2): 293-310.
- Atiya, A. F. 2001. Bankruptcy prediction for credit risk using neural networks: a survey and new results. *IEEE Transactions on Neural Networks* 12 (4): 929-935.
- Baesens, B. 2003. Developing Intelligent Systems for Credit Scoring Using Machine Learning Techniques, Ph.D. Thesis no 180 Faculteit Economische en Toegepaste Economische Wetenschappen, Katholieke Universiteit, Leuven.
- Baesens B, Gestel T V, Stepanova M, Van den Poel D., Vanthienen J 2005. Neural network survival analysis for personal loan data. *Journal of the Operational Research Society* 56 (9): 1089-1098.
- Baesens, B., Gestel, T. V., Viaene, S., Stepanova, M., Suykens, J., Vanthienen, J. 2003. Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring. *Journal of the Operational Research Society* 54 (6): 627-635.
- Baestaens, D-E. 1999. Credit risk modeling strategies: the road to serfdom?. *Intelligent Systems in Accounting, Finance and Management* 8(4): 225-235.
- Bailey, M. 2001. Credit scoring: the principles and practicalities. Kingswood, Bristol: White Box Publishing.
- Bailey, M. 2004. Consumer credit quality: underwriting, scoring, fraud prevention and collections. Kingswood, Bristol: White Box Publishing.
- Banasik, J., Crook, J. 2010 Reject inference in survival analysis by augmentation. *Journal of Operational Research Society* 61 (3): 473-458.
- Banasik J., Crook J. 2007. Reject inference, augmentation, and sample selection. *European Journal of Operational Research* 183 (3): 1582-1594.
- Banasik J. Crook J. 2005. Credit scoring, augmentation, and lean models. *Journal of the Operational Research Society* 56 (9): 1072-1081.
- Banasik, J., Crook, J., Thomas, L. 2001. Scoring by Usage. *Journal of the Operational Research Society* 52 (9): 997-1006.
- Banasik, J., Crook, J., Thomas, L. 2003. Sample Selection Bias in Credit Scoring Models. *Journal of the Operational Research Society* 54 (8): 822-832.
- Behrman, M., Linder, R., Assadi, A. H., Stacey, B. R., Backonja, M. M. 2007. Classification of patients with pain based on neuropathic pain symptoms: Comparison of an artificial neural network against an established scoring system. *European Journal of Pain* 11 (4): 370-376.
- Bellotti, T., Crook, J. 2009. Support vector machines for credit scoring and discovery of significant features. *Expert Systems with Applications* 36 (2/2): 3302-3308.
- Ben-David, A., Frank, E. 2009. Accuracy of machine learning models versus "hand crafted" expert systems – a credit scoring case study. *Expert Systems with Applications* 36 (3/1): 5264-527.

- Bensic, M., Sarlija, N., Zekic-Susac, M. 2005. Modelling small-business credit scoring by using logistic regression, neural networks and decision trees. *Intelligent Systems in Accounting, Finance and Management* 13(3): 133-150.
- Beynon, M. J. 2005. Optimizing object classification under ambiguity/ignorance: application to the credit rating problem. *Intelligent Systems in Accounting, Finance and Management* 13(2): 113-130.
- Bishop, C. M. 1995. *Neural Networks for Pattern Recognition*. New York: Oxford University Press Inc.
- Blochlinger, A., Leippold, M. 2006. Economic Benefit of Powerful Credit Scoring. *Journal of Banking & Finance* 30(3): 851-873.
- Boritz, J. E., Kennedy, D. B. 1995. Effectiveness of neural network types for prediction of business failure. *Expert Systems with Applications* 9 (4): 503–512.
- Boyes, W. J., Hoffman, D. L., Low, S. A. 1989. An Econometric Analysis of the Bank Credit Scoring Problem. *Journal of Econometrics* 40 (1): 3-14.
- Boyle, M., Crook, J. N., Hamilton, R., Thomas, L. C. 1992. Methods for credit scoring applied to slow payers. In *Credit Scoring and Credit Control*, Thomas, L. C., Crook, J. N., Edelman, D. B., eds., Oxford University Press, Oxford, 75-90.
- Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J. 1984. *Classification and Regression Trees*. Belmont: The Wadsworth.
- Cameron, A. C., Trivedi, P. K. 1996. 12 count data models for financial data. *Handbook of Statistics*, 14: 363-391.
- Caouette, J. B., Altman, E. I., Narayanan, P. 1998. *Managing Credit Risk: The Next Great Financial Challenge*. New York: John Wiley & Sons Inc.
- Capon, N. 1982. Credit scoring systems: A critical analysis. *Journal of Marketing* 46 (2): 82-91.
- Carter, D. A., McNulty, J. E. 2005. Deregulation, technology change, and the business-lending performance of large and small banks. *Journal of Banking and Finance* 29 (5): 1113-1130.
- Casu, B., Girardone, C., Molyneux, P. 2006. *Introduction to Banking*. London: Prentice Hall.
- Chandler, G. G., Coffman, J. Y. 1979. A comparative analysis of empirical vs. judgemental credit evaluation. *The Journal of Retail Banking* 1 (2): 15-26.
- Chen, Y., Guo, R-J., Huang, R-L. 2009. Two stages credit evaluation in bank loan appraisal', *Economic Modelling* 26 (1): 63-70.
- Chen, M., Huang, S. 2003. Credit scoring and rejected instances reassigning through evolutionary computation techniques. *Expert Systems with Applications* 24(4): 433-441.
- Chiang, W. K., Zhang, D., Zhou, L. 2006. Predicting and explaining patronage behavior toward web and traditional stores using neural networks: A comparative analysis with logistic regression. *Decision Support Systems* 41 (2): 514–531.

- Chuang, C., Lin, R. 2009. Constructing a reassigning credit scoring model. *Expert Systems with Applications* 36 (2/1): 1685-1694.
- Coffman, J. Y. 1986. The proper role of tree analysis in forecasting the risk behaviour of borrowers. MDS Reports 3, 4, 7 and 9. Management Decision Systems, Atlanta.
- Cramer, J. S. 2004. Scoring bank loans that may go wrong: A case study. *Statistica Neerlandica*, 58 (3): 365-380.
- Crook, J., Edelman D., Thomas, L. 2007. Recent developments in consumer credit risk assessment. *European Journal of Operational Research* 183 (3): 1447-1465.
- Crook, J. N. 1996. Credit scoring: An overview. Working paper series No. 96/13, British Association, Festival of Science. University of Birmingham, The University of Edinburgh.
- Dasgupta, C. G., Dispensa, G. S., Ghose, S. 1994. Comparing the predictive performance of a neural network model with some traditional market response models. *International Journal of Forecasting* 10 (2): 235–244.
- Delen, D., Walker, G., Kadam, A. 2005. Predicting breast cancer survivability: A comparison of three data mining methods. *Artificial Intelligence in Medicine* 34 (2): 113–127.
- Desai, V. S., Conway, D. G., Crook, J. N., Overstreet, G. A. 1997. Credit scoring models in the credit union environment using neural networks and genetic algorithms. *IMA Journal of Mathematics Applied in Business and Industry* 8 (4): 323-3463.
- Desai, V. S., Crook, J. N., Overstreet, G. A. 1996. A Comparison of Neural Networks and Linear Scoring Models in the Credit Union Environment. *European Journal of Operational Research* 95 (1): 24-37.
- Deschaine, L., Francone, F. 2008. Comparison of Discipulus™ Linear Genetic Programming Soft-ware with Support Vector Machines, Classification Trees, Neural Networks and Human Experts, White Paper. Available at: <http://www.rmltech.com/> (Accessed: 10 June 2008).
- DeYoung, R. Frame, W. S., Glennon, D., McMillen, D. P., Nigro, P. 2008. Commercial lending distance and historically underserved area. *Journal of Economics and Business* 60 (1-2): 149-164.
- Dimla, D. E., Lister, P. M. 2000. On-line metal cutting tool condition monitoring. II: tool-state classification using multi-Layer perceptron neural networks. *International Journal of Machine Tools & Manufacture* 40 (5): 769-781.
- Duliba, K. 1991. Contrasting neural nets with regression in predicting performance. In *Proceedings of the Twenty-Fourth Annual Hawaii International Conference on System Sciences* 4, 163-170, IEEE Press, Alamitos, CA.
- Durand, D. 1941. Risk Elements in Consumer Instalment Financing, *Studies in Consumer Instalment Financing*. New York: National Bureau of Economic Research.
- Dutta, S., Shekhar, S., Wong, W. Y. 1994. Decision support in non-conservative domains: Generalization with neural networks. *Decision Support Systems*, 11 (5): 527–544.

- Dvir, D., Ben-Davidb, A., Sadehb, A., Shenhar, A. J. 2006. Critical managerial factors affecting defense projects success: A comparison between neural network and regression analysis. *Engineering Applications of Artificial Intelligence* 19 (5): 535–543.
- Eisenbeis, RA. 1977. Pitfalls in the application of discriminant analysis in business, finance, and economics. *The Journal of Finance* XXXII (3): 875-900.
- Eisenbeis, R. A. 1978. Problems in Applying Discriminant Analysis in Credit Scoring Models. *Journal of Banking and Finance* 2 (3): 205-219.
- Elliott, R., Filinkov, A. 2008. A self tuning model for risk estimation. *Expert Systems with Applications* 34 (3): 1692-1697.
- Emel, A., Oral, M., Reisman, A., Yolalan, R. 2003. A credit scoring approach for the commercial banking sector. *Socio-Economic Planning Sciences* 37 (2): 103-123.
- Etemadi, H., Rostamy, A., Dehkordi, H. 2009. A genetic programming model for bankruptcy prediction: Empirical evidence from Iran', *Expert Systems with Applications* 36 (2/2): 3199-3207.
- Falbo, P. 1991. Credit-scoring by enlarged discriminant model. *Omega*, 19 (4): 35-54.
- Feelders, A. J. 2000. Credit scoring and reject inference with mixture models. *Intelligent Systems in Accounting, Finance and Management* 9(1): 1-8.
- Feng, C.-X., Wang, X. 2002. Digitizing uncertainty modeling for reverse engineering applications: Regression versus neural networks. *Journal of Intelligent Manufacturing* 13 (3): 189–199.
- Finney, P. J. 1952. *Probit Analysis*, Cambridge, MA: Cambridge University Press.
- Fisher, R. A. 1936. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics* 7 (2): 179-188.
- Fletcher, D., Goss, E. 1993. Forecasting with neural networks: An application using bankruptcy data. *Information and Management* 24 (3): 159–167.
- Foglia, A., Laviola, S., Reedtz, P. 1998. Multiple banking relationships and the fragility of corporate borrowers. *Journal of Banking and Finance* 22 (10-11): 1441-1456.
- Frame, W., Padhi, M., Woosley, L. 2004. Credit scoring and the availability of small business credit in low-and moderate-income areas. *The Financial review* 39 (1): 35-54.
- Frame, W., Srinivasan, A., Woosley, L. 2001. The effect of credit scoring on small-business lending. *Journal of Money, Credit and Banking* 33 (3): 815-825.
- Freund, R. J., William, W. J. 1998. *Regression analysis: Statistical modeling of a response variable*. San Diego: Academic Press.
- Fritz, S., Hosemann, D. 2000. Restructuring the credit process: behaviour scoring for german corporates. *Intelligent Systems in Accounting, Finance and Management* 9(1): 9-21.
- Gately, E. 1996. *Neural Networks for Financial Forecasting: Top Techniques for Designing and Applying the Latest Trading Systems*. New York: John Wiley & Sons, Inc.
- Glen J. 2001. Classification accuracy in discriminant analysis: a mixed integer programming approach. *Journal of the Operational Research Society* 52 (3): 328-339.

- Greene, W. 1998. Sample Selection in Credit-Scoring Models. *Japan and the World Economy* 10 (3): 299-316.
- Golberg, D. E. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*, reading. Boston, MA: Addison-Wesley.
- Grablowsky, B. J., Talley, W. K. 1981. Probit and discriminant functions for classifying credit applicants: a comparison. *Journal of Economic and Business*, 33 (3): 254-261.
- Guillen, M., Artis, M. 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.
- Gup, B. E., Kolari, J. W. 2005. *Commercial Banking: The management of risk*. Alabama: John Wiley & Sons, Inc.
- Hand, D. J. 1981. *Discrimination and Classification*. New York: John Wiley & Sons Inc.
- Hand, D. J. 1997. *Construction and assessment of classification rules*. Chichester: John Wiley & Sons Inc.
- Hand, D. J., Henley, W. E. 1997. Statistical Classification Methods in Consumer Credit Scoring: A Review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 160 (3): 523-541.
- Hand, D. J., Jacka, S. D. 1998. *Statistics in Finance, Arnold Applications of Statistics*: London.
- Hand, D. J., Oliver, J. J., Lunn, A. D. 1998. Discriminant analysis when the classes arise from a continuum. *Pattern Recognition* 31 (5): 641-650.
- Hand, D. J., Sohn, S. Y., Kim, Y. 2005. Optimal bipartite scorecards. *Expert Systems with Applications* 29(3): 684-690.
- Hardgrave, B. C., Wilson, R. L., Walstrom, K. A. 1994. Predicting graduate student success: A comparison of neural networks and traditional techniques. *Computers and Operations Research* 29 (30): 49-263.
- Haughwout, A., Peach, R., Tracy, J. 2008. Juvenile delinquent mortgages: bad credit or bad economy? *Journal of Urban Economics* 64 (2): 246-257.
- Haykin, S. 1994. *Neural networks: A comprehensive foundation*. London: Prentice Hall, Inc.
- Heiat, A. 2002. Comparison of artificial neural network and regression models for estimating software development effort. *Information and Software Technology* 44 (15): 911-922.
- Heffernan, S. 2005. *Modern Banking*. Chichester, West Sussex: John Wiley & Sons, Inc.
- Henley, W. E. 1995. *Statistical aspects of credit scoring*. Ph.D. Thesis, The Open University, Milton Keynes.
- Henley, W. E., Hand, D. J. 1996. A k-nearest-neighbour classifier for assessing consumer credit risk. *The Statistician* 45 (1): 77-95.
- Heuson, A., Passmore, W., Sparks, R. 2001. Credit scoring and mortgage securitization: implications for mortgage rates and credit availability. *Journal of Real Estate Finance and Economics* 23 (3): 337-363.

- Hill, T., Remus, W. 1994. Neural network models for intelligent support of managerial decision making. *Decision Support Systems* 11 (5): 449–459.
- Hoffmann, F., Baesens, B., Mues, C., Gestel, T. V., Vanthienen, J. 2007. Inferring descriptive and approximate fuzzy rules for credit scoring using evolutionary algorithms. *European Journal of Operational Research* 177 (1): 540-555.
- Hosmer, D. W., Lemeshow, S. 1989. *Applied Logistic Regression*. New York: John Wiley & Sons, Inc.
- Hsieh, N-C 2004. An integrated data mining and behavioral scoring model for analysing bank customers. *Expert Systems with Applications* 27 (4): 623-633.
- Hu, Y. 2008. Incorporating a non-additive decision making method into multi-layer neural networks and its application to financial distress analysis. *Knowledge-Based Systems* 21 (5): 383-390.
- Hu, Y-C., Ansell, J. 2007. Measuring retail company performance using credit scoring techniques. *European Journal of Operational Research* 183 (3): 1595-1606.
- Huang, C., Chen, M., Wang, C. 2007. Credit scoring with a data mining approach based on support vector machines. *Expert Systems with Applications* 33 (4): 847-856.
- Huang, J., Tzeng, G., Ong, C. 2006. Two-stage genetic programming (2SGP) for the credit scoring model. *Applied Mathematics and Computation* 174 (2): 1039-1053.
- Ignizio, J. P., Soltys, J. R. 1996. An ontogenic neural network for bankruptcy classification. *IMA Journal of Mathematics Applied in Business & industry* 7 (4): 313-325.
- Irwin, G. W., Warwick, K., Hunt, K. J. 1995. *Neural networks applications in control*. London: The Institution of Electronic Engineers.
- Jo, H., Han, I., Lee, H. 1997. Bankruptcy prediction using case-based reasoning, neural network and discriminant analysis. *Expert Systems with Applications* 13 (2): 97–108.
- Kay, J. W., Titterton, D. M. 1999. *Statistics and neural networks: advanced at the interface*. New York: Oxford University Press.
- Kim, Y. S., Sohn, S. Y. 2004. Managing Loan Customers Using Misclassification Patterns of Credit Scoring Model. *Expert Systems with Applications* 26 (4): 567-573.
- Koza, J. R. 1992. *Genetic Programming On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: MIT Press.
- Koza, J. R. 1994. *Genetic Programming II Automation Discovery of Reusable Programs*. Cambridge, MA: MIT Press.
- Krishnaswamy, M., Krishnan, P. 2002. Nozzle wear rate prediction using regression and neural network. *Biosystems Engineering* 82 (1): 53–64.
- Kumar, A., Motwani, J. 1999. Reengineering the lending procedure for small businesses: A case study. *Work Study* 48 (1): 6-12.
- Kumar, A., Rao, V. R., Soni, H. 1995. An empirical comparison of neural network and logistic regression models. *Marketing Letters* 6 (4): 251–263.
- Kumra, R., Stein, R., Assersohn, I. 2006. Assessing a knowledge-based approach to commercial loan underwriting. *Expert Systems with Applications* 30 (3): 507-518.

- Landajo, M., Andres, J. D., Lorca, P. 2007. Robust neural modelling for the cross-sectional analysis of accounting information. *European Journal of Operational Research* 177 (2): 1232–1252.
- Laha, A. 2007. Building contextual classifiers by integrating fuzzy rule based classification technique and k-nn method for credit scoring. *Advanced Engineering Informatics* 21 (3): 281-291.
- Lee, K., Booth, D., Alam, P. 2005. A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms. *Expert Systems with Applications* 29 (1): 1–16.
- Lee, T., Chen, I. 2005. A Two-Stage Hybrid Credit Scoring Model Using Artificial Neural Networks and Multivariate Adaptive Regression Splines. *Expert Systems with Applications* 28 (4): 743-752.
- Lee, T., Chiu, C. Lu, C., Chen, I. 2002. Credit Scoring Using the Hybrid Neural Discriminant Technique. *Expert Systems with Applications* 23 (3): 245-254.
- Lee, T. H., Jung, S. 2000. Forecasting creditworthiness: Logistic vs. artificial neural net. *The Journal of Business Forecasting Methods and Systems* 18 (4): 28–30.
- Lenard, M. J., Alam, P., Madey, G. R. 1995. The application of neural networks and a qualitative response model to the auditor's going concern uncertainty decision. *Decision Sciences* 26 (2): 209–227.
- Lensberg, T. Eilifsen, A., McKee, T. 2006. Bankruptcy theory development and classification via genetic programming. *European Journal of Operational Research* 169 (2): 766-697.
- Leonard, K. J. 1996. Information Systems and Benchmarking in the Credit Scoring Industry. *Benchmarking for Quality Management & technology* 3 (1): 38-44.
- Leonard, K. J. 1995. The development of credit scoring quality measures for consumer credit application. *International Journal of Quality & Reliability Management* 12 (4): 79-85.
- Leonard, K. J. 1993. Detecting credit card fraud using expert systems. *Computers & Industrial Engineering* 25 (1-4): 103-106.
- Leonard, K. J. 1992. Credit scoring models for the evaluation of small-business loan applications. *IMA Journal of Mathematics Applied in Business & Industry* 4 (1): 89-95.
- Leshno, M., Spector, Y. 1996. Neural network prediction analysis: The bankruptcy case. *Neurocomputing* 10 (2): 125–147.
- Lewis, E. M. 1992. *An Introduction to Credit Scoring*. California: Fair, Isaac & Co., Inc.
- Liang, Q. 2003. Corporate Financial Distress Diagnosis in China: Empirical Analysis Using Credit Scoring Models. *Hitotsubashi Journal of Commerce and Management* 38 (1): 13-28.
- Lim, M. K., Sohn, S. Y. 2007. Cluster-Based Dynamic Scoring Model. *Expert Systems with Applications* 32 (2): 427-431.
- Limsombunchai, V., Gan, C., Lee, M. 2005. An analysis of credit scoring for agricultural loans in Thailand. *American Journal of Applied Sciences* 2 (8): 1198–1205.

- 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.
- Lovie, A. D. 1987. The bootstrapped – Lessons for the acceptance of intellectual technology. *Applied Ergonomics* 18 (3): 201-206.
- Lucas, A. 1992. Updating scorecards: removing the mystique. In *Credit Scoring and Credit Control*, Thomas, L. C., Crook, J. N., Edelman, D. B., eds., Oxford University Press, Oxford, 180-197.
- Maddala, G. S. 2001. *Introduction to Econometrics*. Chichester: John Wiley & Sons Inc.
- Malhotra, R., Malhotra, D. K. 2003. Evaluating consumer loans using Neural Networks. *Omega the International Journal of Management Science* 31 (2): 83-96.
- Masters, T. 1995. *Advanced Algorithms for Neural Networks: AC++ Sourcebook*. New York: John Wiley & Sons, Inc.
- Mays, E. 2001. *Handbook of Credit Scoring*. Chicago: Glenlake Publishing Company, Ltd.
- Mays, E. 2004. The Rule of Credit Scores in Consumer Lending. In E. Mays, *Credit Scoring for Risk Managers: The Handbook for Lenders*. (3-12). Australia: Thomson South-Western.
- McKee, T., Lensberg, T. 2002. Genetic programming and rough sets: A hybrid approach to bankruptcy classification. *European Journal of Operational Research* 138 (2): 436-451.
- Min, J. H., Lee, Y-C 2008. A practical approach to credit scoring. *Expert Systems with Applications* 35 (4): 1762-1770.
- Min, J. H., Jeong, C. 2009. A binary classification method for bankruptcy prediction. *Expert Systems with Applications* 36(3): 5256-5263.
- Mukkamala, S., Vieira, A., Sung, A. 2008. Model selection and feature ranking for financial distress classification. Available at: <http://www.rmltech.com/> (Accessed: 10 June 2008).
- Myers, J. H., Forgy, E. W. 1963. The development of numerical credit evaluation systems. *Journal of the American Statistical Association* 58 (303): 799-806.
- Nanni, L., Lumini, A. 2009. An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring. *Expert Systems with Applications* 36 (2/2): 3028-3033.
- Nakamura, E. 2005. Inflation forecasting using a neural network. *Economics Letters* 86 (3): 373-378.
- Nath, R., Rajagopalan, B., Ryker, R. 1997. Determining the saliency of input variables in neural network classifiers. *Computers and Operations Researches* 24 (8): 767–773.
- Nelson, M. M., Illingworth, W. T. 1990. *A Practical Guide to Neural Nets*. New York: Addison Wesley.
- Nguyen, T., Malley, R., Inkelis, S. H., Kuppermann, N. 2002. Comparison of prediction models for adverse outcome in pediatric meningococcal disease using artificial neural

- network and logistic regression analyses. *Journal of Clinical Epidemiology* 55 (7): 687–695.
- Nikolopoulos, K., Goodwin, P., Patelis, A., Assimakopoulos, V. 2007. Forecasting with cue information: A comparison of multiple regression with alternative forecasting approaches. *European Journal of Operational Research* 180 (1): 354–368.
- Nunez-Letamendia, L. 2002. Trading Systems Designed by Genetic Algorithms. *Managerial Finance* 28 (8): 87-106.
- Ong, C., Huang, J., Tzeng, G. 2005. Building Credit Scoring Models Using Genetic Programming. *Expert Systems with Applications* 29 (1): 41-47.
- Orgler, Y. E. 1971. Evaluation of Bank Consumer Loans with Credit Scoring Models. *Journal of Bank Research* 2 (1): 31-37.
- Orgler, Y. E. 1970. A credit scoring model for commercial loans. *Journal of Money, Credit and Banking* II (4): 435-445.
- Ottensbacher, K. J., Smith, P. M., Illig, S. B., Linn, R. T., Mancuso, M., Granger, C. V. 2004. Comparison of logistic regression and neural network analysis applied to predicting living setting after hip fracture. *Annals of Epidemiology* 14 (8): 551–559.
- Paleologo, G., Elisseeff, A., Antonini, G. 2010 Subagging for credit scoring models. *European Journal of Operational Research* 201 (2): 490-499.
- Palisade Corporation. 2005. Neural Tools: Neural Networks Add-In for Microsoft Excel. Version 1.0. New York: Palisade Corporation.
- Paliwal, M., Kumar, U. A. 2009. Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications* 36 (1): 2-17.
- Pendharkar, P. C. 2005. A threshold-varying artificial neural network approach for classification and its application to bankruptcy prediction problem. *Computers and Operations Research* 32 (10): 2561–2582.
- Pindyck, R. S., Rubinfeld, D. L. 1997. *Econometric Models and Economic Forecasts*. McGraw-Hill/Irwin.
- Quah, J. T.S., Sriganesh, M. 2008. Real-time credit card fraud detection using computational intelligence. *Expert Systems with Applications* 35 (4): 1721-1732.
- Raiffa, H., Schlaifer, R. 1961. *Applied Statistical Decision Theory*. Boston: Harvard University Press.
- Reed, R. D., Marks, R. J. 1999. *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks*. London: The MIT Press.
- Rosenberg, E., Gleit, A. 1994. Quantitative methods in credit management: a survey. *Operations Research* 42 (4): 589-613.
- Salchenberger, L. M., Cinar, E. M., Lash, N. A. 1992. Neural networks: A new tool for predicting thrift failures. *Decision Sciences* 23 (4): 899–916.
- Sarlija, N., Bensic M., Bohacek Z. 2004. Multinomial Model in Consumer Credit Scoring, 10th International Conference on Operational Research. Trogir: Croatia.

- Sarlija, N., Bensic, M., Zekic-Susac, M. 2009. Comparison procedure of predicting the time to default in behavioural scoring. *Expert Systems with Applications*, 36 (5): 8778-8788.
- Seow, H., Thomas, L. C. 2006. Using Adaptive Learning in Credit Scoring to Estimate Take-Up Probability Distribution. *European Journal of Operational Research* 173 (3): 880-892.
- Shang, J. S., Lin, Y. E., Goetz, A. M. 2000. Diagnosis of MRSA with neural networks and logistic regression approach. *Health Care Management Science* 3 (4): 287–297.
- Siddiqi, N. 2006. *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*. New Jersey: John Wiley & Sons, Inc.
- Sinha AP, Richardson MA. 1996. A Case-Based Reasoning System for Indirect Bank Lending. *Intelligent Systems in Accounting, Finance and Management* 5 (4): 229-240.
- Smith, A. E., Mason, A. K. 1997. Cost estimation predictive modeling: regression versus neural network. *The Engineering Economist* 42 (2): 137–161.
- Somers, M., Whittaker, J. 2007. Quantile regression for modelling distributions of profit and loss. *European Journal of Operational Research* 183 (3): 1477-1487.
- Song, J. H., Venkatesh, S. S., Conant, E. A., Arger, P. H., Sehgal, S. M. 2005. Comparative analysis of logistic regression and artificial neural network for computer-aided diagnosis of breast masses. *Academic Radiology* 12 (4): 487–495.
- Spear, N. A., Leis, M. 1997. Artificial neural networks and the accounting method choice in the oil and gas industry. *Accounting Management and Information Technology* 7 (3): 169–181.
- Steenackers, A., Goovaerts, M. J. 1989. A Credit Scoring Model for Personal Loans. *Insurance: Mathematics and Economics* 8 (8): 31-34.
- Stefanowski, J., Wilk, S. 2001. Evaluating business credit risk by means of approach-integrating decision rules and case-based learning. *Intelligent Systems in Accounting, Finance and Management* 10(2): 97-114.
- Sullivan, A. C. 1981. Consumer Finance. In E. I. Altman, *Financial Handbook* (9.3-9.27), New York: John Wiley & Sons.
- Sustersic, M., Mramor, D., Zupan J. 2009. Consumer credit scoring models with limited data. *Expert Systems with Applications* 36 (3): 4736-4744.
- Tam, K. Y., Kiang, M. Y. 1992. Managerial applications of neural networks: The case of bank failure predictions. *Management Science* 38 (7): 926–947.
- Teller, A., Veloso, M. 2000. Internal reinforcement in a connectionist genetic programming approach. *Artificial Intelligence* 120 (2): 165-198.
- Thanh Dinh, T-H., Kleimeier, S. 2007. A credit scoring model for Vietnam's retail banking market. *International Review of Financial Analysis* 16 (5): 471-495.
- Thieme, R. J., Song, M., Calantone, R. J. 2000. Artificial neural network decision support systems for new product development project selection. *Journal of Marketing research* 37 (4): 499–507.

- Thomas, L. C. 1998. Methodologies for Classifying Applicants for Credit' in Hand, D. J. & Jacka, S. D. (eds.), *Statistics in Finance* (83-103), London: Arnold.
- Thomas, L. C. 2000. A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International Journal of Forecasting* 16 (2): 149-172.
- Thomas, L. C., Edelman, D. B., Crook, J. N. 2004. *Readings in Credit Scoring: recent developments, advances, and aims*. New York: Oxford University Press.
- Thomas, L. C., Edelman, D. B., Crook, L. N. 2002. *Credit Scoring and Its Applications*. Philadelphia: Society for Industrial and Applied Mathematics.
- Thompson, P. 1998. Bank Lending and the Environment: Policies and Opportunities. *International Journal of Bank Marketing* 16 (6): 243-252.
- Trinkle, B. S., Baldwin, A. A. 2007. Interpretable credit model development via artificial neural networks. *Intelligent Systems in Accounting, Finance and Management* 15(3-4): 123-147.
- Trippi, R. R., Turban E. 1993. *Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real-World Performance*. Chicago: IRWIN.
- Tsai, C., Wu, J. 2008. Using neural networks ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications* 34 (4): 2639-2649.
- Tsaih, R., Liu, Y., Liu, W., Lien Y. 2004. Credit scoring system for small business loans. *Decision Support Systems* 38 (1): 91-99.
- Usha, A. K. 2005. Comparison of neural networks and regression analysis: A new insight. *Expert Systems with Applications* 29 (2): 424-430.
- Van Gestel, T., Baesens, B., Suykens, J. A.K., Van den Poel, D., Baestaens, D-E., Willekens, M. 2006. Bayesian kernel based classification for financial distress detection. *European Journal of Operational Research* 172 (3): 979-1003.
- Verstraeten, G., Van den Poel, D. 2005. The impact of sample bias on consumer credit scoring performance and profitability. *Journal of the Operational Research Society* 56 (8): 981-992.
- Walczak, S., Sincich, T. 1999. A comparative analysis of regression and neural networks for university admissions. *Information Sciences* 119 (1-2): 1-20.
- Warner, B., Misra, M. 1996. Understanding neural networks as statistical tools. *The American Statistician* 50 (4): 284-293.
- West, D. 2000. Neural Network Credit Scoring Models. *Computers & Operations Research* 27 (11-12): 1131-1152.
- West, D., Dellana, S., Qian, J. 2005. Neural network ensemble strategies for financial decision applications. *Computers & Operations Research* 32 (10): 2543-2559.
- Xia, Y., Liu, B., Wang, S., Lai, K. K. 2000. A model for portfolio selection with order of expected returns. *Computers & Operations Research* 27 (5): 409-422.
- Yang, Y. 2007. Adaptive credit scoring with kernel learning methods. *European Journal of Operational Research* 183 (3): 1521-1536.

- Yang, Z., Wang, Y., Bai, Y., Zhang, X. 2004. Measuring Scorecard Performance. Computational Science-ICCS LNCS 3039, 900-906.
- Yesilnacar, E., Topal, T. 2005. Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). Engineering Geology 79 (3–4), 251–266.
- Yim J., Mitchell H. 2005. Comparison of country risk models: hybrid neural networks, logit models, discriminant analysis and cluster techniques. Expert Systems with Applications 28 (1): 137-148.
- Yoon, Y., Swales, G., Jr., Margavio, T. M. 1993. A comparison of discriminant analysis versus artificial neural networks. The Journal of the Operational Research Society 44 (1): 51–60.
- Yu, L., Wang S., Lai, K. 2009. An intelligent-agent-based fuzzy group decision making model for financial multicriteria decision support: the case of credit scoring. European Journal of Operational Research 195 (3): 942-959.
- Zekic-Susac, M., Sarlija, N., Bensic, M. 2004. Small Business Credit Scoring: A Comparison of Logistic Regression, Neural Networks, and Decision Tree Models. 26th International Conference on Information Technology Interfaces. Croatia.
- Zhang, Y., Bhattacharyya, S. 2004. Genetic programming in classifying large-scale data: an ensemble method. Information Sciences 163 (1-3) : 85-101.
- Zhang, G., Hu, M. Y., Patuwo, B. E., Indro, D. C. 1999. Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. European Journal of Operational Research 116 (1): 16–32.

TABLES

Table 1: Classification results for different scoring models (%)

Model	Total correct classification	Correct classification of good	Correct classification of bad	The percentage of bad accepted into the good group
Discriminant analysis	65.4	62.2	78.0	8.1
Linear regression model	55.1	47.0	87.5	6.2
Probit model	71.9	76.4	54.1	13.1
Poisson model	62.4	57.7	81.8	7.3
Negative binomial II model	63.3	58.9	80.6	7.6
Two step procedure	64.9	61.1	79.8	7.6

Source: Guillen & Artis (1992, p. 9), adapted.

Table 2: Comparison of the bad risk rates using different scoring techniques

Scoring technique	Bad risk rate (%)
K-NN (any D)	43.09
K-NN (D = 0)	43.25
Logistic regression	43.30
Linear regression	43.36
Decision tree	43.77

Notation: K-NN = k-nearest-neighbour. It is a standard technique in pattern recognition and non-parametric/non-linear statistics, to credit scoring problems. Source: Henley & Hand (1996, p. 91).

Table 3: Comparison of the ACC rates using different scoring techniques

Scoring technique	ACC rate (%)		
	Sample ₁	Sample ₂	Sample ₃
Multiple Discriminant Analysis	78.05	78.84	79.40
Logistic Regression	82.33	81.46	82.09
Probabilistic Neural Network	85.90	87.32	87.64
Multi-layer Feed-forward Neural Network	84.07	84.47	87.48

Sample₁ denote 67/33% for the training/testing sets; Sample₂ denote 80/20% for the training/testing sets; Sample₃ denote 90/10% training/testing sets. Source: Abdou & Pointon (2009, p. 401), modified.

Table 4: Statistically significant differences, and credit scoring errors: comparing models and credit data

	German credit	Australian credit
Superior models	MOE	MOE
	RBF	RBF
	MLP	MLP
	Logistic reg.	Logistic reg.
		LDA
Inferior models		K nearest neighbor
	LVQ	LVQ
	FAR	FAR
	LDA	Kernel density
	K nearest neighbor	CART
	Kernel density	
	CART	

Notation: MOE = Mixture of Experts; RBF = Radial Basis Function; MLP = Multi-Layer Perceptron; LVQ = Learning Vector Quantization; FAR = Fuzzy Adaptive Resonance, all of which are neural networks models; and LDA = Linear Discriminant Analysis. Source: West (2000, p. 1145).

Table 5: Comparing classification results for different scoring models

Scoring Model	Correctly classified results	
	Testing	Overall
Weight of Evidence Model	52.16	54.99
Probit Analysis	82.69	81.93
Genetic Programming – Best Programme (GP _p)	82.93	83.28
Genetic Programming – Best Team (GP _t)	83.89	85.82

Source: Abdou (2009c, pp. 11411-11412), modified.

Table 6: A comparison of percentage correctly classified from published research

	Boyle et al. (1992)	Desai et al. (1997)	West (2000) ¹	Lee et al. (2002)	Malhotra & Malhotra (2003)	Baesens (2003) ³	Ong et al. (2005) ⁴
Linear regression or LDA	77.5	66.5	79.3	71.4	69.3	79.3	80.8
Logistic regression		67.3	81.8	73.5		79.3	
Decision tree	75.0		77.0			77.0	78.4
Math programming	74.7					79.0	
Neural nets		66.4	82.6	73.7 (77.0) ²	72.0	79.4	81.7
Genetic programming							82.8
K-nearest neighbours			76.7			78.2	
Support vector machines						79.7	

¹Figures are an average across two data-sets. ²Hybrid LDA and NN. ³Figures are an average of eight data-sets. ⁴Figures are an average over two data-sets. Source: Crook et al. (2007, p. 1457).

Table 7: Frequently used performance evaluation criteria

Error measure	No. of papers
Confusion matrix	36
MSE/RMSE	16
MAE	7
Mean error	2
R^2 /Adj R^2	2
Sensitivity, specificity analysis using ROC curve	7
Goodness of fit test (Calibration)	3
Discrimination (C-statistic/AUC)	5

Notation: MSE = Mean Square Error; RMSE = Root Mean Square Error; MAE = Mean Absolute Error; ROC curve = Receiver Operating Characteristics curve; AUC = Area Under the ROC Curve, all of which are performance evaluation criteria. Source: Paliwal & Kumar (2009, p. 14), adapted.

FIGURES

Figure 1: The Receiver Operating Characteristics (ROC) curve



The proportion of bad cases classified as bad (vertical axis) against the proportion of good cases classified as bad (horizontal axis) at all cut-off score values can be represented by the ROC curve. If the proportion of bad cases classified as bad equal to the proportion of good cases classified as bad, in this case there is no separation at all and the distribution are identical, the ROC curve will lie over the slanting straight line; therefore the proportion of the area below the ROC curve which is above the slanting line can be used as a measure of the separation yielded by a scoring model. Source: Crook et al. (2007, p. 1450), adapted.