Editorial

Credit Scoring

Academic and commercial research into credit scoring has mushroomed over the last decade. The range of issues considered, the range of countries in which academic research on the topic is developing and the range of journals now containing credit scoring papers are all increasing exponentially. However the largest academic conference on credit scoring is held biennially in the UK, by the Credit Research Centre, University of Edinburgh, and it is from the eighth Edinburgh conference in 2003 that almost all of the papers in this Special Issue are drawn. This Issue contains some particularly exciting papers that are of both practical and academic importance though, of course, these papers are not intended to cover all the areas in which research is being pursued.

There have been significant developments in the consumer credit environment in the last few years, which have hastened and broadened research in the area. The first paper by Thomas *et al* assesses past research and sets out their view of an agenda for future work. In particular, they argue that the objectives of credit models have widened from merely default risk prediction to, for example, profitability scoring, consumer scoring and risk-based pricing. Some of these changes have been encouraged by the new Basel Capital Accord, which allows banks to use their own models, provided they can satisfy the regulators, to determine the minimum amount of capital they should retain to cover unexpected loan portfolio losses. Lenders may classify loans into sub-portfolios of similar risk, which together with the methods to be used to predict losses required by the regulators, for example loss given default averaged over several years, are a further major subject of research. Indeed, the risk models incorporated in the Accord come from corporate risk modelling and a new area of research involves assessing and adapting these to the retail loan sector. Of course, searching for the most accurate classification algorithm will remain a perennial topic of research and research into reject inference is ongoing. The authors predict a further research area of the optimal characteristics of loans which will appeal to the wishes of borrowers. Many of these issues are reflected in the papers included in this special issue.

The next four papers specifically concern aspects of profit assessment rather than the traditional concern of predicting the probability of default. The second paper, by Beling *et al*, uses the efficient frontier concept of Oliver and Wells[1] to derive conditions under which, if a lender has two scorecards

available to score the same applicant, one card should be used rather than the other. They show that if scorecard A has a receiver operating curve, ROC curve, always above that of scorecard B (ROC dominance) this is equivalent to the expected profit from A dominating that of B. On the other hand, if one scorecard does not dominate the other, then maximizing the expected profit depends on the slope of a line tangent to the ROCs of both scorecards together with the slope of a line indicating constant profits (the isoprofit line). The slope of the ROC is the *a priori* prior odds divided by the loss:gain ratio. Given these relative slopes, one then examines the conditions in which the expected profit—expected volume curve implied by each scorecard dominates the other.

In the third paper, Oliver and Keeney develop a similar theme by putting together the utility function of the lender with that of a potential borrower. The objectives of the lender are to maximize a utility function, which has as arguments both expected profits and market share. The authors use a utility function that is, in fact, found in what economists call discrete choice theory. The potential borrower's utility depends negatively on interest rate (price) and positively on credit line (quality). The lower the former and the higher the latter, the greater the chance the potential customer will accept a loan offer and hence the greater the market share of the lender. This then implies that there is a profit maximizing choice of interest rate to offer. Offering lower rates increases the probability of customer acceptance and hence the volume, but the drop in profit per customer more than compensates for this. Offering higher rates makes the customers more profitable, but the drop in the volume of acceptances reduces the overall profit. The optimal rate for a lender to offer depends on his trade-off between the two goals of profit and market share. There is no doubt that this approach will spawn considerable further research.

The next two papers use survival analysis to model a consumer's purchase behaviour over time, first by use of a store card and second by the purchase of a financial product. Most store cards are issued just before the time when they are first used. An important question relating to the profits that can be earned by the lender is when will the consumer make a second purchase? The paper by Andreeva *et al* shows that using new information about a customer's behaviour, as this information becomes available, allows a lender to better predict the chance that a card will be used to make a second purchase in the next time period. Specifically, adding the nature of the first product bought, its price and payment date to information available only at the time of application

improved predictive performance. Including subsequent information, which is observed at later points in time, especially the amount available to spend when the card was issued and, subsequently, amount to spend in two adjacent periods, further improved predictive and explanatory performance. The paper by Thomas *et al* models the time to a second purchase of a financial product (any of an investment product, pension product and protection product or life policy) from an insurance company. They find that both in a competing-risk and single-risk approach, age and socio-demographic categories significantly affect the type of product purchased. But the main message is that to understand time to purchase behaviour of financial products one needs to incorporate macroeconomic factors into a model. The specific variables the authors experimented with were asset prices, consumer confidence, earnings and interest rates, though the latter had no separate effect.

The remaining papers are advances in how the probability of default may be predicted more accurately. Moffatt wishes to model the volume arrears on an account after a given exposure period. He distinguishes between those accounts that never default and so never have arrears above zero from those which potentially may default. Now one could model arrears using a standard Tobit model. But this assumes that the parameters in the model predicting whether an account defaults are the same as those in the model which explains the magnitude of arrears. Moffatt shows, by using Cragg's double hurdle model, that separately modelling whether an account is a potential defaulter and the volume of arrears enables one to tease out the effects of different variables in understanding the latter. The results are quite dramatic showing that not only do different applicant variables explain the two processes, but some variables that appear in both processes have totally different impacts in the two processes. For example, males are less likely to ever default, but will have a greater volume of arrears if they do. Since the Basel II requirements involve the estimation of loss given default, this methodology could have very wide application.

Banasik and Crook examine a potential problem that may affect all credit risk models: the nature of bias which may result from estimating an application default model using only the performance of previously accepted applicants. Unlike most other researchers, they use application data where (virtually) every applicant was granted credit. They empirically examine whether any such bias varies with the number of variables in the model and whether any improvement in performance, when a standard method of reject inference is used, also varies with the size of the model. They find that the deterioration in performance when an accept-only model is used is greater when the model is larger and when the previous model was used with a higher cut-off so that only the very best applicants were accepted. The contribution of augmentation is small at all model sizes and at high original cut-offs augmentation can actually reduce predictive performance at all model sizes.

Schebesch *et al* and Baesens *et al* examine the accuracy of relatively new classification methods. While an increasing number of papers have assessed the accuracy of different classification algorithms when applied to credit assessment data, the recent development of support vector machines has received very little attention in the published literature on credit scoring. The paper by Schebesch and Stecking adds to this limited literature by giving an accessible explanation of SVMs including an explanation of linear and non-linear SVMs and how one might deal with the case where a very large number of support vectors are found. They use SVMs to give an idea of their predictive performance when applied to credit application data distinguishing between typical cases that are easy to predict and critical cases where it is more difficult to predict class membership. Baesens *et al* examine a different algorithm, neural networks, but in a new application: that of parameterizing consumer default and early repayment survival models. Using data on a personal loan the authors compare the predictive performance of a network, which has as many output nodes as time periods considered. Each output is coded according to whether a case is good, in default or censored. Each model is compared with a Cox regression and a logistic regression for predicting default in the first 12 months and, separately, the next 12 months. They find that the network was more accurate than logistic regression, but just as accurate as Cox regression when predicting early repayment. When predicting default all three methods gave indistinguishable results, unless the defaults were deliberately oversampled. This suggests that neural nets do not lead to significant improvements in modelling survival data to predict defaults or early repayment compared with existing statistical approaches. However, there may be greater improvements if other network architectures are employed and, if so, one still has to decide why the more flexible nature of the neural net approach cannot be more exploited.

The paper by Liu and Schumann continues the theme of machine learning in credit scoring by examining four data mining techniques for the selection of features—predictor variables—for application of default-scoring models. The experiments are carried out in conjunction with the use of four classification algorithms: decision trees, neural networks, logistic regression and nearest neighbours. The authors conclude that the consistency based and wrapper feature selection methods reduce the features without reducing accuracy more rapidly than other selection methods. When features selected by these methods are incorporated in the classification algorithm, only the k-nearest neighbour algorithm improves accuracy. The feature selection methods also differ considerably in terms of speed and number of features selected—both relevant criteria for method selection.

Of course to decide which algorithm performs best, given a specific data set, requires a measure of performance. In the

next paper Hand argues that many conventional measures of the performance of an application scorecard, such as the Gini, the Kolmogorov-Smirnov statistic and the information statistic, all use information which is not relevant and ignore information which is. If a decision is to be made by comparing a score with a cut-off, and the costs of misclassification are the same for an accepted bad and for a rejected good, then the distribution of scores is not needed and may indeed be misleading. The appropriate measure, according to Hand, is the proportion of applicants who are classified as good but who turn out to be bad. When a model is to be used to select customers for specific action, for example increase a credit limit, the same weakness applies and Hand proposes another criterion which should be used in this case. When the costs of misclassification differ according to the errors, then a transformation of the Gini is appropriate. The simple message is: choose your measure of

performance according to the problem at hand. It will be interesting to see whether the industry and academic researchers heed this warning.

JN Crook
Credit Research Centre,
University of Edinburgh
DE Edelman
Caledonia Credit Consultancy and
Dataplex Data Services
LC Thomas
School of Management,
University of Southampton

References

1 Oliver RM and Wells E (2001). Efficient frontier cutoff policies in credit portfolios. *J Opl Res Soc* **52**: 1025–1033.