Opinion piece

ROC analysis in credit scoring and credit judgment

Received (in revised form): 1st November, 2005

T. P. Hutchinson

is on the staff of the Centre for Automotive Safety Research, University of Adelaide. Besides traffic safety, he also has wide interests in applications of statistical methods. For further information see casr.adelaide.edu.au/paul

Abstract Relative operating characteristic (ROC) analysis is useful in the context of (global) judgments about whether or not to grant credit. But scepticism is expressed about its relevance when detailed scores of creditworthiness are available.

INTRODUCTION

Desirable customers have a distribution of credit scores. Undesirable customers also have a distribution of credit scores. Assuming the credit scoring system is a good one, the lower the credit (risk) score, the greater is the ratio of desirable to undesirable customers. (This was how the scale was defined by Hoogenberg and auf dem Brinke.¹) A cut-off point can be chosen, with customers having a higher score being refused service. A high cut-off point will mean that a high proportion of desirable customers are accepted along with quite a high proportion of undesirable customers. A low cut-off point will mean that both these proportions are lower. Too high a cut-off and the costs associated with undesirable customers will be high; too low a cut-off and desirable customers are turned away. If one plots the proportion of desirable customers turned down versus the proportion of undesirable customers turned down (the different points on the graph corresponding to different cut-offs), then the result is called a relative operating characteristic (ROC).²

ROC analysis has had many other applications, some of which are described by Hutchinson^{3,4} and Swets *et al.*⁵ In the finance/economics field, a graph equivalent to the ROC was used in 1980 by Ohlson.⁶ It is not obvious why and how the recent explosion in interest came about; as well as those referred to by Hoogenberg and auf dem Brinke, papers by Wodon⁷ and Karim and Siegel⁸ seem to be early examples.

Signal detection theory (SDT) is also pertinent in this field: the idea is that the task is to perceive a distinction between a desirable customer and an undesirable customer, analogous to detecting a faint visual or auditory signal in the presence of background noise. In this context, the ROC dates back to the 1950s.

In marketing and advertising, there has been a degree of familiarity with SDT at least since the mid-1980s.^{9,10}

The ROC is an important concept. The trade-off between two types of mistake is quite a simple idea, but its implications and ramifications are not apparent unless the subject is presented systematically, as by Hoogenberg and auf dem Brinke.¹¹ Their article concentrated

T. P. Hutchinson
Centre for Automotive
Safety Research,
University of Adelaide,
Adelaide,
South Australia 5005,
Australia
Tel: +61 8 8303 5997;
Fax: +61 8 8232 4995;
e-mail:
paul@casr.adelaide.edu.au

on laxity versus strictness in decision making. A feature of ROC analysis is that this issue is distinguished from that of the separation between the two distributions — of credit scores of desirable and undesirable customers, for example. One credit scoring model might discriminate between desirable and undesirable customers more effectively than another model. If too many classification mistakes are noticed, it can be important to know whether this stems from an inappropriate degree of strictness or from discrimination being too poor, as it is typically easier to change the degree of strictness than the effectiveness of discrimination.

While the ROC is an important concept for credit *judgment*, this paper questions whether it adds much to credit *scoring*.

WHEN IS ROC ANALYSIS USEFUL?

Despite what Hoogenberg and auf dem Brinke said,¹² and despite the many related articles (just type “credit score ROC” into Google to find some of them), the author is sceptical about whether ROC analysis does add much to credit scoring. As already mentioned, the classic context for the ROC is that of a human attempting to perceive a faint light or sound in the presence of noise. Credit scoring would be analogous if the true status of each potential customer were binary (desirable and undesirable), the classification crude (only two or a few categories, such as high, medium, low) and the decision binary (accept or reject); however credit scoring is usually different.

In the first place, it is natural to view a customer being undesirable or desirable (defaulting or not defaulting) as an outcome, which might possibly be predicted from the information about the customer. ROC analysis seems to turn

this around, taking undesirable versus desirable as the initial distinction and then comparing the credit scores. Secondly, more information than just the credit score often available, eg, the information that went into calculating the credit score. Hoogenberg and auf dem Brinke themselves touch on this point when they note the possibility that profit may depend upon credit score¹³. Thirdly, many different credit scores may occur, not only two or three categories. Finally, the decision itself need not always be binary (accept or reject): intermediate categories — such as ‘seek further information’ or ‘accept but charge a higher price’ may be available.

Where there are many different credit scores occurring, common statistical techniques may be a more natural choice than ROC analysis.

- First, suppose that there are only two categories of customer, desirable and undesirable. These may be compared pictorially (for example, using histograms or cumulative distributions or boxplots). They may also be compared using appropriate summary statistics: means and standard deviations if the shapes of the distributions are approximately normal (although the choice of summary statistics may be difficult when the distributions are skewed).
- Secondly, consider the situation where there are more than two true categories of how good a customer is (eg, how profitable), the latter can be plotted versus the credit score to show how successful a predictor it is.

It thus seems that several common statistical techniques can be used in credit scoring.

None of the above is a knockout blow against ROC analysis, and I have already expressed the opinion that the

concept is an important one. Nevertheless, I do think it would be useful if Hoogenberg and auf dem Brinke could pinpoint the reasons why they prefer ROC analysis over either prediction of undesirability by (for example) logistic regression, or (if they insist on starting from the division into undesirable or desirable) several common techniques for comparing distributions.

Many other researchers have used ROC in a similar manner to Hoogenberg and auf dem Brinke:

- Various authors in the field of finance have followed the same general approach.
- It is very common to find papers on medical diagnosis that compare test results (eg, concentrations of some chemical) of patients with and without a particular disease giving an ROC in full detail.
- Examples in various other fields may be found, eg, in ecology.¹⁴ The second edition of the well-known textbook on logistic regression by Hosmer and Lemeshow¹⁵ includes a description of how to use ROC to assess the discriminating ability of a model.

So people do employ the ROC in this way, but the author is still not clear what distinctive contribution ROC analysis makes, and in what situations. Two come to mind:

- 1 If the question is specifically, 'If I accept wrongly rejecting x per cent of desirable customers, then what percentage of undesirable customers will I correctly identify?', ROC is exactly what one wants.
- 2 Drawing the ROC for two models (such as a credit scoring model based upon many characteristics and one based upon only one or two) on a

single graph enables them to be easily compared.

ROC analysis is really useful when common statistical techniques cannot be used, for example, when someone deciding whether or not to grant credit may have an unfavourable intuition about an applicant without being able to articulate a reason. The question then arises whether the credit officer's intuition has any validity or not. A decision based on subjective judgment would be investigated within the ROC framework by comparing the proportions of bad outcomes when the judgment was favourable and when it was unfavourable. The credit officer may be willing to describe the strength of a subjective feeling (none, slight, definite suspicion), and the validity of this could be investigated in a similar fashion. In the context of the creditworthiness of borrowers from banks, Lehmann has reported on improved accuracy from taking into account subjective judgments as well as hard facts, like financial ratios.^{16,17} Other decisions where subjective judgment plays a big part (perhaps, whether or not to hire someone or whether or not to purchase something) could be investigated in the same way.

Acknowledgments

The Centre for Automotive Safety, University of Adelaide, receives core funding from the Motor Accident Commission (South Australia) and the Department for Transport, Energy and Infrastructure (South Australia). The views expressed in this paper are those of the author, not necessarily of the University of Adelaide nor the sponsoring organisations.

References

- 1 Hoogenberg, M. and auf dem Brinke, S. (2004) 'Decision making in an uncertain environment: An application of ROC analysis for credit scoring in the mobile telephone market', *Journal of Database Marketing and Customer Strategy Management*, Vol. 11, pp. 241–254.
- 2 Hoogenberg, M. and auf dem Brinke, S. (2004) *Ibid.*

- 3 Hutchinson, T. P. (1981) 'A review of some unusual applications of signal detection theory', *Quality and Quantity*, Vol. 15, pp. 71–98.
- 4 Hutchinson, T. P. (1985) 'Presenting one probability distribution as a function of another — Some applications', *American Journal of Mathematical and Management Sciences*, Vol. 5, pp. 103–123.
- 5 Swets, J. A., Dawes, R. M. and Monahan, J. (2000) 'Psychological science can improve diagnostic decisions', *Psychological Science in the Public Interest*, Vol. 1, pp. 1–26.
- 6 Ohlson, J. A. (1980) 'Financial ratios and the probabilistic prediction of bankruptcy', *Journal of Accounting Research*, Vol. 18, pp. 109–131.
- 7 Wodon, Q. T. (1997) 'Targeting the poor using ROC curves', *World Development*, Vol. 25, pp. 2083–2092.
- 8 Karim, K. E. and Siegel, P. H. (1998) 'A signal detection theory approach to analyzing the efficiency and effectiveness of auditing to detect management fraud', *Managerial Auditing Journal*, Vol. 13, pp. 367–375.
- 9 Hutchinson, J. W. and Zenor, M. (1986) 'Product familiarity and the strengths of brand-attribute associations: A signal detection theory approach', *Advances in Consumer Research*, Vol. 13, pp. 450–453.
- 10 Singh, S. N. and Churchill, G. A. (1986) 'Using the theory of signal detection to improve ad recognition testing', *Journal of Marketing Research*, Vol. 23, pp. 327–336.
- 11 Hoogenberg, M. and auf dem Brinke, S. (2004) *op cit.*
- 12 Hoogenberg, M. and auf dem Brinke, S. (2004) *ibid.*
- 13 Hoogenberg, M. and auf dem Brinke, S. (2004) *ibid.*, p. 253.
- 14 Pearce, J. and Ferrier, S. (2000) 'Evaluating the predictive performance of habitat models developed using logistic regression', *Ecological Modelling*, Vol. 133, pp. 225–245.
- 15 Hosmer, D. W. and Lemeshow, S. (2000) 'Applied Logistic Regression. Second Edition', Wiley, New York, NY.
- 16 Lehmann, B. (2003) 'Is it worth the while? The relevance of qualitative information in credit rating', Working Paper, Center of Finance and Econometrics, University of Konstanz, Germany.
- 17 Lehmann, B. (2004) 'How good is "good"? — Managing subjective information in credit ratings', Working Paper, Center of Finance and Econometrics, University of Konstanz, Germany.