Linearising the ROC and detecting asymmetry: A credit rating example

The relative operating characteristic (ROC) was originally used as a signal of possible default, and is usually illustrated as a curve on a linear axis. **PAUL HUTCHINSON** shows that ROC also works if it is linearised

T. P. Hutchinson
Centre for Automotive Safety Research
University of Adelaide, South Australia

The relative operating characteristic (ROC) has become a familiar idea in studies of credit rating. Suppose some form of credit (risk) rating is available, with low scores corresponding to high risk. Consider the fraction *HR* (hit rate) of defaulters who have a score less than some cut-off, and the fraction *FAR* (false alarm rate) of non-defaulters who have a score less than the cut-off.

The ROC is the line traced out by plotting *HR* versus *FAR*, as the cut-off score is varied. The general shape is illustrated in Figure 1. For more on this, see, for example, Engelmann et al. (2003). In some descriptions, such as Sobehart and Keenan (2001), the convention followed is that high scores correspond to high risk.

Besides defaulting on credit, other commercial contexts for ROC analysis include bankruptcy of businesses, people becoming bad debtors, sovereign nations defaulting, fraudulent versus genuine vehicle insurance claims, fraudulent misuse versus normal use of mobile phones, computer network operation under attack or normal conditions, printed pages incorrectly or correctly read by machine, personal biological measurements incorrectly or correctly processed by biometric technology.

In all of these cases, the task can be viewed as the attempted detection of a "signal" (such as future default) in conditions of uncertainty. Indeed, one of the origins of ROC analysis was in what is termed signal detection theory in the psychology of perception. For applications outside of commerce of attempting to detect a signal that distinguishes one group from another, see Hutchinson (1981).

It surprises me that in credit rating contexts the ROC is shown as a curve on linear axes, and there is no attempt to transform the axes so that the line becomes straight, or approximately so. Such transformation is common in psychology and other fields. The purpose of this paper is to demonstrate linearisation with a credit rating example.

METHOD

In what follows, *z(p)* is the standard normal deviate corresponding to a cumulative probability *p*. That is, for a distribution that is normal with mean = 0 and standard deviation = 1, *z* is the quantity such that the probability of being less than it is *p*. (Example: when *p* is 0.975, *z* is 1.96.)

- A plot of *z(HR)* versus *z(FAR)* will usually be approximately a straight line.
- The reason behind using *z* is that if there exists a monotonic transformation of credit scores such that the distributions are normal for defaulters and for non-defaulters, the result will be exactly a straight line. (This transformation is not necessarily

anything simple. There is no need to know what it is.)

• The straight line has a slope of 1 if these two transformed distributions have equal standard deviations. Furthermore, in this case, the ROC (on linear axes) is symmetric about the line $HR = 1 - FAR$.

• If the standard deviations are unequal, the straight line on transformed axes has some other slope, and the untransformed ROC is asymmetric.

If software to compute z from a probability is not conveniently available, the transformation $ln(p/(1-p))$ (where *ln* means natural logarithm) might be used as an alternative. This is appropriate for logistic distributions, rather than normal distributions, of the defaulters and non-defaulters.

The reasons for considering linearisation to be helpful are common ones when considering visual presentation of data: as a mental aid in appreciating the relationship, for ease of use of the equation for interpolation (and, more riskily, for extrapolation), to summarise the relationship in terms of the slope and intercept, and (if a model's prediction of a straight line is seen to be closely correct) to make it plausible that the model is close enough to correct for other purposes also.

RESULTS

In Table 2 of Krämer and Güttler (2003) and Table 2 of Güttler (2005), data are given on the ratings of 1927 companies (mostly industrial firms and financial institutions from the US), 209 of which defaulted.

Ratings by both Moody's and by Standard & Poor's were available, each on a 17-point scale. Some general characteristics of the dataset are the following:

(1) The proportion of defaulters is high, around 11% (it is a four-year cumulative default rate);

(2) The median rating (Moody's) is B3 for defaulters and Baa1 for non-defaulters;

(3) There is no evidence that the top seven categories (Aaa to A3, some 43% of the companies) differ in respect of probability of default: this is close to zero for all of them.

Indeed, $z(HR)$ is undefined for the highest categories (i.e. is at infinity), and thus does not appear on the linearised ROC. As to the ROC, on linear axes, this is shown as Figure 1 of Güttler (2005).

Figure 2 shows the ROC linearised as suggested above, i.e., as $z(HR)$ versus $z(FAR)$. Data for Moody's and Standard & Poor's are shown by different symbols. The chief features of Figure 2 are as follows:

(1) Each relationship (for Moody's and Standard & Poor's) is approximately linear;

(2) The two relationships are almost indistinguishable, even though it is typically easier to see a difference when both are nearly linear than on the original, untransformed, axes;

(3) Their slopes are not 1. (Consequently, the ROCs are not symmetrical.)

The slope and intercept of the straight lines are interpretable as follows. Imagine that the credit score for non-defaulters has been transformed so that the distribution is normal, with a mean of 0 and a standard deviation of 1. Suppose that transformed credit score for defaulters now has a mean of μ and a standard deviation of σ. Then at a cut-off C, $z(FAR)$ is C and:

FIGURE 1

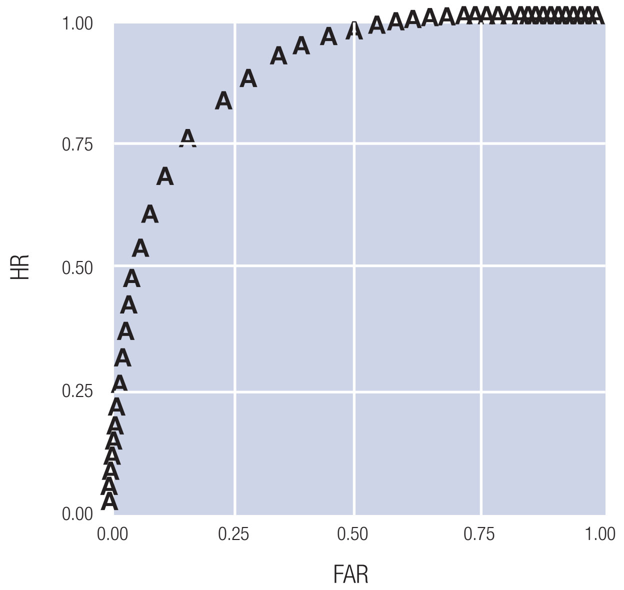

FIGURE 1: General shape of the ROC: increasing from (0,0) to (1,1), HR exceeds FAR, and may or may not be symmetric about HR = 1 - FAR.

FIGURE 2

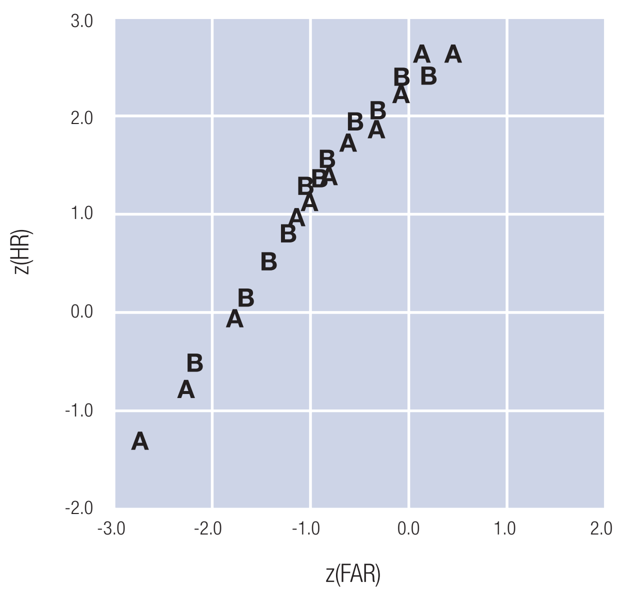

FIGURE 2: Approximate linearity of z(HR) plotted against z(FAR). **B**s represent data from Moody's, **A**s represent data from Standard & Poor's. (Data from Table 2 of Krämer and Güttler, 2003, and Table 2 of Güttler, 2005.)

z(HR) is $(C - \mu)/\sigma$. Consequently, $z(HR) = (1/\sigma).z(FAR) - (\mu/\sigma)$.

The degree of discrimination between defaulters and non-defaulters is often summarised as the area under the curve (AUC). A by-product of the analysis here is that it becomes clear that if σ is 1, μ can be used as a measure of how different the distributions are (and some people may consider this to be more easily appreciated than the AUC); and that there may be comprehensible features of the ROC (such as σ) that cannot be captured if one statistic alone is selected.

The slopes in Figure 2 are greater than 1 (about 1.3). Thus the standard deviation for defaulters is smaller than for non-defaulters, for both Moody's and Standard & Poor's. (A limitation on this conclusion is that the scale is defined so that the distribution for non-defaulters is normal; it does not necessarily apply in the case of other definitions, for example, if scores of 1, 2, 3, … 17 are used for the categories. Actually, for this dataset, use of scores 1, 2, 3, … 17 also results in a smaller standard deviation for defaulters.)

CONCLUSION

ROC analysis takes two groups (such as defaulters and non-defaulters) and looks back to see how they compared in terms of some quantitative or ordinal variable such as credit rating. It thus complements the process of starting with credit rating (or other variables) and predicting default or non-default as an outcome. Different ways of looking at data have different strengths and weaknesses.

There has been some recent attention given to asymmetry of the ROC. Liu (2002) notes the possibility of *HR* obtained by one method being higher than for another method for some range of *FAR*, but being lower for a different range, and uses asymmetric ROCs in the illustrations reinforcing this point. Appendix I of Bemmann (2005) includes consideration of the possibilities that *HR* is a power function of *FAR*, and that *1 – HR* is a power function of *1 – FAR*. Both are asymmetric. Incidentally, transformation to straight lines is simply by taking logarithms. Blöchlinger and Leippold, (2004/2005) discuss some consequences, in terms of the appropriate pricing of loans, of differences in shape between ROCs having the same area under the curve.

Whether the ROC is symmetric about *HR = 1 – FAR* or not, in my view a relationship is easier to grasp if it is a straight line, or nearly so. Transformation of an ROC to near linearity deserves to be better known.

Acknowledgements

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 or the sponsoring organisations.

References

Bemmann, M. (2005), "Improving the comparability of insolvency predictions". *Discussion paper in Economics No. 08/05*, Faculty of Business Management and Economics, Dresden University of Technology.

Blöchlinger, A. and Leippold, M. (2004/2005), "Economic benefit of powerful credit scoring". *Working paper No. 216*, National Centre of Competence in Research Financial Valuation and Risk Management, University of Zurich.

Engelmann, B., Hayden, E. and Tasche, D. (2003), "Measuring the discriminative power of rating systems". *Discussion paper (Series 2: Banking and Financial Supervision) No. 01/2003*, Deutsche Bundesbank, Frankfurt am Main.

Güttler, A. (2005), "Using a bootstrap approach to rate the raters". *Working paper No. 132*, Department of Finance, Goethe University, Frankfurt am Main.

Hutchinson, T. P. (1981), "A review of some unusual applications of signal detection theory". *Quality and Quantity*, 15, 71–98.

Krämer, W. and Güttler, A. (2003), "Comparing the accuracy of default predictions in the rating industry: The case of Moody's vs. S&P". *Discussion paper No. 23/03*, Sonderforschungsbereich (Collaborative Research Centre) 475, University of Dortmund.

Liu, Y. (2002), "The evaluation of classification models for credit scoring". *Arbeitsbericht Nr. 02/2002*, Institut für Wirtschaftsinformatik, Georg-August-Universität Göttingen.

Sobehart, J. and Keenan, S. (2001), "Measuring default accurately". *Risk*, 14, Credit Risk Special Report S31–S33. J