Credit scoring and the value of data
Author: Murray Bailey

There are two principles about the use of data for scorecard development. Firstly, the better the data, the better the resultant model. Secondly, if it is legally available, evaluate it for use in the scorecard.

There are three types of data that should be considered for inclusion in application scorecard development:

- application form data
- credit reference information
- existing account information

Application data is taken at the time of application and should include both accepted and rejected applications. Rejects are necessary to fairly represent the complete 'through-the-door' population, since it is to this group that the scorecard will be applied.

Any additional information disclosed or uncovered following investigation by underwriters should also be included. If the underwriters discover application information is inaccurate, both pieces of information are relevant. Wherever possible, avoid overwriting details; errors and falsification can be just as predictive and sometimes more illuminating than accurate information.

The second category of data is all the bureau information at the time of the application or decision. This is ideally captured at the point of application. However, credit reference agencies typically provide a service whereby data is recreated retrospectively.

The existing account data arises where a previous relationship has existed and contains performance information more detailed than that held by the credit reference agency. Own account information may be available real-time, whereas bureau data can be up to two months old.

Grouping characteristics
Data available for scorecard development may range from 30 to 200 or more variables (or characteristics). So, how does the developer select the most powerful ones for inclusion in the final scorecard?

Before we look at this, we should consider two further classifications of the data. Characteristics are either continuous or discrete (categorical). Examples of continuous variables are: age, time at address and credit turnover. Examples of discrete variables are: residential status, marital status and worst arrears status.

A key driver of characteristic interpretation is the number of accounts that can be analysed. In most application scorecard developments, the number of bad accounts is limited and this becomes the main determinant for banding or grouping the attributes. The first stage for continuous variables is often to band them into manageable groups, for example converting time at address to years at address.

The next stage is to combine attributes or groups of attributes into ones that are statistically significant. This is a critical stage of the development and is time consuming. Poor grouping at this stage will result in a weaker scorecard.

Developers either compare bad rates by eye or use the student-t or F test to identify significance. Common sense must always play a part as must legal requirements. The aim is to maximise the discriminatory power of the characteristic - the ability to distinguish between good and bad applicants - without compromising the validity of the data.

Characteristic power
An alternative to the bad rate of attributes is the weight of evidence (WoE). WoEs

are effectively the 'raw scores' of an attribute, reflective of an additive probability. The WoE is the natural logarithm of the information odds, the equation for which is displayed in equation 1.

Information odds are the ratio of goods to bads where the overall population risk has been removed by comparing equal numbers or percentages of the goods and bads. In box 1 $\%Good_a$ is the percentage of the goods with attribute 'a' and $\%Bad_a$ is the percentage of the bads with attribute 'a'.

Equation 1: The weight of evidence equation for attribute 'a'

$$WoE_a = Ln(\%Good_a/\%Bad_a)$$

Since WoE_a reflects the discrimination for attribute 'a' of a characteristic, combining all WoEs provides a measure of the power of the whole characteristic. This is called the Information Value (IV). IV measures the separation of the goods and bads by the characteristic, assuming it to be the only variable in the scorecard.

Box 2 is the equation for IV and table 1 is an example of an Information Value calculation where the IV is the sum of the WoE for each attribute after it is weighted by the difference in proportions of goods and bads.

Equation 2: The Information Value equation for characteristic 'c'

$$\text{Information Value}_c = \Sigma\ (\%Good_a - \%Bad_a) \times WoE_a$$

Table 1 Example Information Value table for Age

Attribute	Information Odds	Weight of Evidence	%Good - %Bad	I.V. contribution
< 25	0.571	-0.560	-15%	0.084
25 - 34	1.000	0.000	0%	0.000
35 +	1.428	0.357	15%	0.054
			Information Value:	0.138

Trade-off between acceptance and risk

Whilst IV provides a statistical test of significance, it does not reflect the impact of the scorecard on the business decision; the trade off between acceptance rate and the bad rate.

Figure 1 is a trade-off chart for a scorecard used for credit card applications. As the acceptance rate increases, so does the bad rate. Hence, the scorecard is discriminating and the degree of discrimination is reflected by the gradient of the curve.

If the business has a cut-off that delivers an acceptance rate of 66%, the bad rate is predicted to be 1.9%. This same chart can be used for assessing the impact and contribution of characteristics.

The most common form of scorecard development today is 'step-wise' regression. Characteristics are stepped into the model one at a time to avoid inclusion of unnecessary or weak variables. Characteristics can also be 'stepped out' of the development, thereby providing a result that can directly reflect the contribution of the characteristic. For example, at the same acceptance rate we could compared the bad rate before and after removal.

If the 'stepping out' process is used during development, the other point scores will change to compensate for correlation. However, simple removal without

redevelopment can also provide a quick guide to the maximum impact of missing characteristics.

Case study

For the purposes of confidentiality, I shall not use the real names of the banks concerned. When two banks merged - West and North Bank - no one realised the implications for the decision systems would be so great.

Both banks had well-established scoring systems, however the similarity ended there. The systems and platforms for the systems were disparate. In addition, West Bank had developed complex models that relied more on bureau data and current account performance than application information.

Atypically, West Bank had developed a complex bespoke credit reference link. This unique link allowed West Bank to create their own bureau characteristics rather than rely on those generated by the bureau. North Bank had a standard link with a different bureau.

After two years of running in parallel, the inevitable decision was made to have a single, enterprise system. The West Bank team assumed it would be their more advanced solution that would survive.

The decision went the other way. The North Bank platform was newer and more widely in operation, so the application scoring system and bureau link would be theirs.

In response, the West Bank team proposed the enhancement of the newer system to meet their needs. This was rejected. The systems team were involved in the “Year 2000” changes and there was no internal resource. Due to the same issue, all external development had been frozen.

The West Bank team were instructed to redevelop their scorecards, based on the information available to North's system.

Trade-off charts

Since good scorecard development is time consuming, whilst they began the redevelopment, the West Bank team calculated the impact of the missing data. They looked at the change in acceptance rate for the same bad rate and the change in bad rate for the same acceptance rate. This was called the “data gap”.

West Bank had three main portfolios: credit cards, personal loans and overdrafts. The trade-off charts for the three main scorecards are shown in figures 2 to 4. The dotted line is the original trade-off chart. The solid line is the trade-off chart assuming the loss of characteristics using North's system – without redevelopment.

For the credit card, personal loan and overdraft portfolios, the acceptance rates were: 66%, 75% and 74% respectively. The data gap in acceptance rates (at the same bad rates) were 5%, 15% and 10% respectively. The data gap in bad rates (at the same acceptance rate) were 0.13%, 0.24% and 0.07% respectively.

The average deterioration in bad rate due to loss of the characteristics was 10% (an average data gap of 0.14%). The biggest impact was clearly on the personal loan. The greater impact on the personal loan new business is clearly seen in figure 3, which shows a greater separation of the two lines.

This was to be expected due to the high level of cross-sell of the personal loan to existing customers. The credit card scorecard was least impacted because the population was more generic with a high proportion of applicants new to the bank.

Loss impact

The Credit Manager presented the charts to the management committee and pointed out the impact on write-offs. If the acceptance rates were maintained, the loss of the characteristics would mean an additional write-off per year of approximately £2.7 million.

The figures by product are shown in table 2. Whilst the personal loan portfolio had the greatest data gap, the credit card portfolio would suffer higher loss impact due to the higher volume of new business.

Table 2 Annual credit loss impact by product

Product	New Business	Average Write-off	Bad rate	Bad rate data gap	Write-off Impact
Credit cards	500,000	£2,300	1.70%	0.13%	£1.50m
Personal loans	200,000	£2,600	0.74%	0.24%	£1.15m
Overdrafts	300,000	£1,000	0.18%	0.07%	£0.21m
Total	900,000	£2,066	1.04%	0.15%	£2.66m

The Credit Manager also pointed out that the manner in which West Bank had created the characteristics. The unique variables provided a competitive advantage that was demonstrated by the impact on losses but also customer value and cross-sell opportunity.

The priority associated with the scorecards suddenly changed. Clearly the Year 2000 system enhancements took priority, but after that came the enhancement of North's system to handle the bureau link and new scoring characteristics.

The West Bank team also completed the redevelopment of the scorecards with and without the jeopardised characteristics. The revised scorecards, including the unique characteristics, achieved improvements over and above those shown in figures 2 to 4.

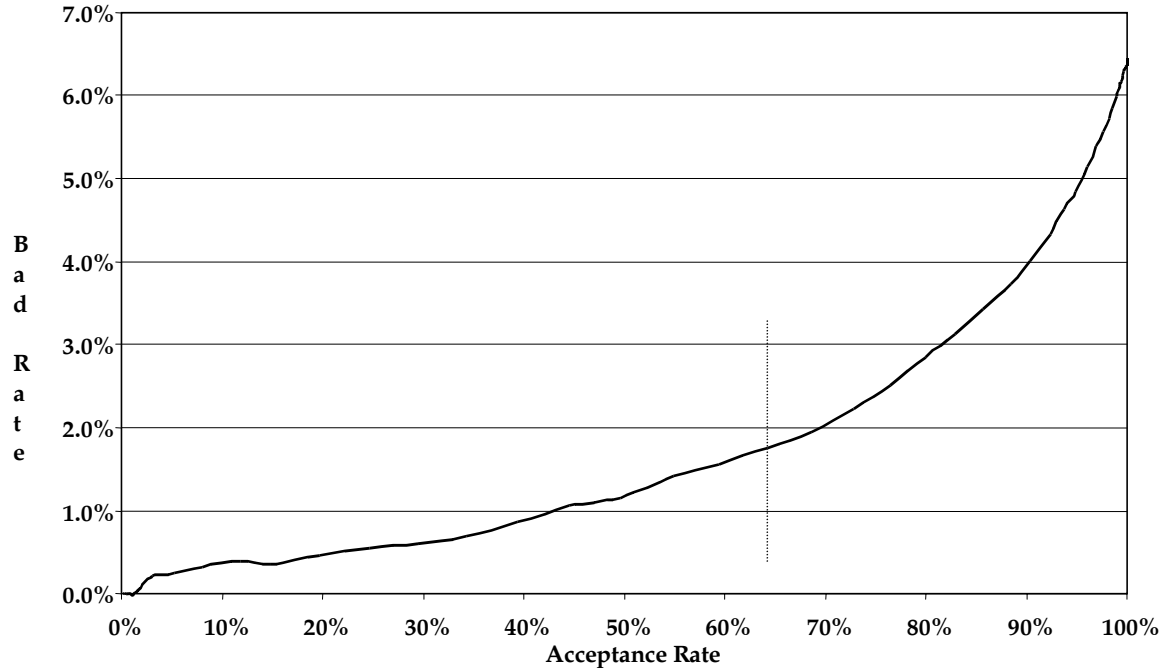

Figure 1 Example trade off chart

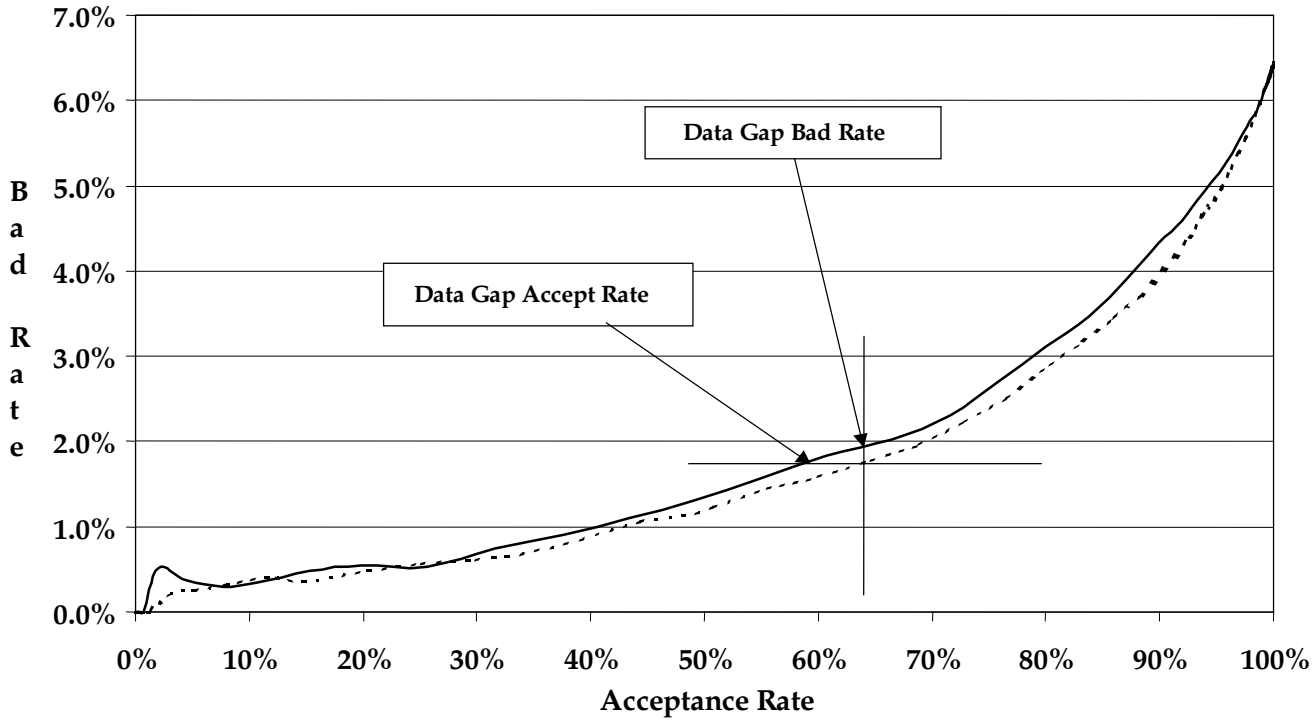

Figure 2 Trade off charts for credit cards

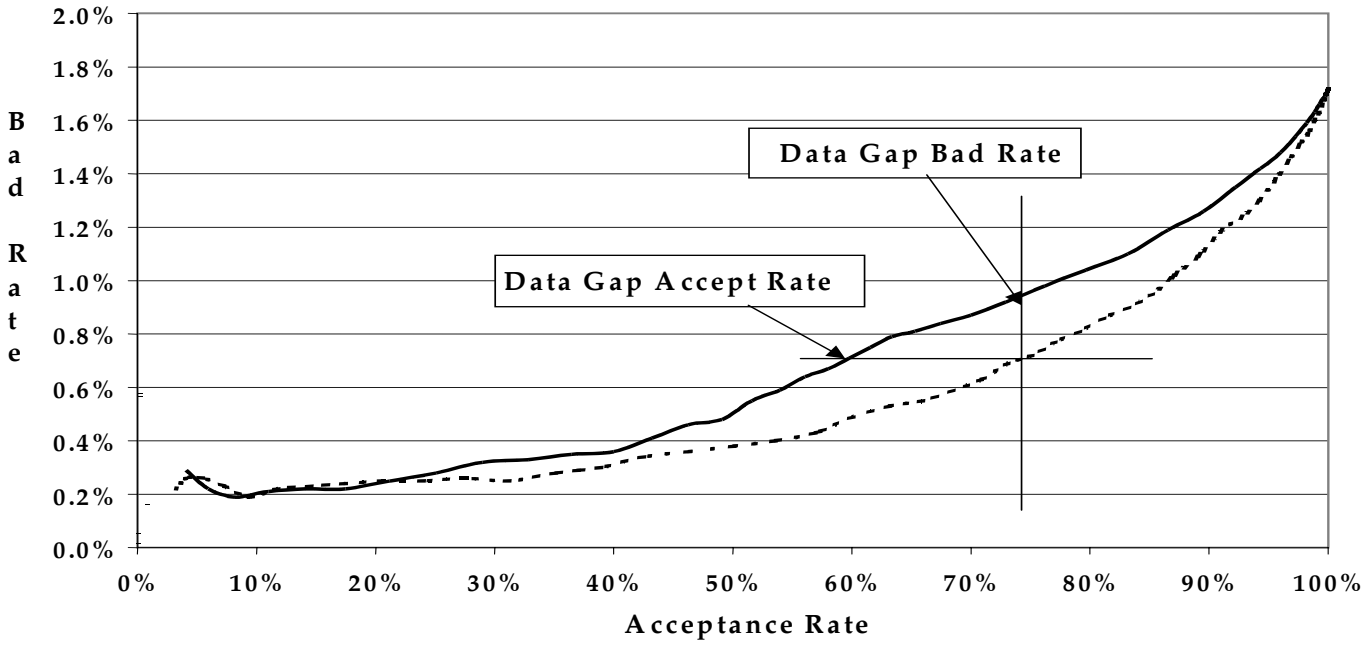

Figure 3 Trade off charts for personal loans

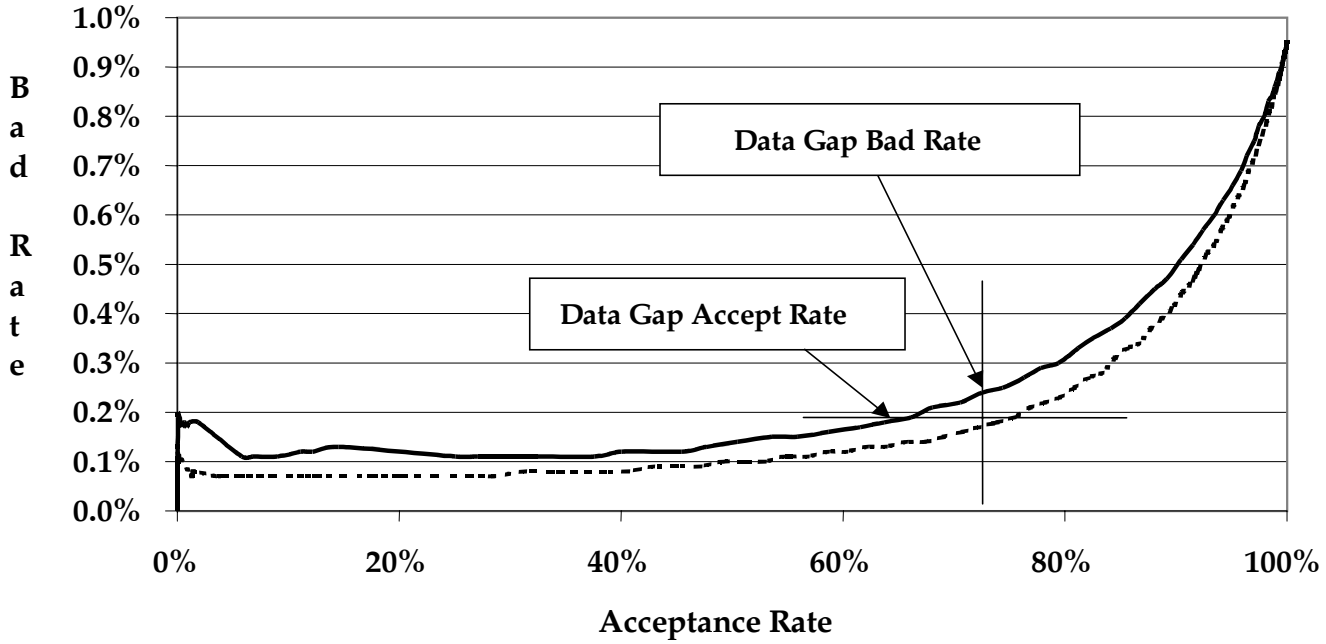

Figure 4 Trade off charts for overdrafts

Author:

Murray Bailey is managing director of Windsor CME
Email: murray.bailey@windsorcme.co.uk