The Good, the Bad and the Ugly

By Murray Bailey

What is Bad?
Fundamentally we would want to define Bad as a customer with whom we would not have done business had we known about his performance in advance.

Ideally we would classify the write-offs as the Bads, but it often (hopefully) takes too long to wait for all the write-offs. If a developer classifies all write-offs over the life of a portfolio as Bad and all non-write-offs as Good, the Gini of the resulting scorecard is usually very high. The problem is that the oldest write-off might be so old that the characteristics that were predictive when it was booked are less reflective of the current business. There is also an issue of whether the business will be happy to approve customers who spend time in collections but do not progress to write-off. There may also be a problem with timing of the write-off if different term loans are available in a portfolio. The result will be a variable performance period.

Figure 5.1 illustrates the ideal single Performance period where the proportion ever bad flattens. Sometimes this doesn't occur over a reasonable time frame and the developer is forced to select a shorter period. It is good practice to check the proportion becoming Bad over this period to ensure it is in excess of 50% of the accounts that are ever Bad.

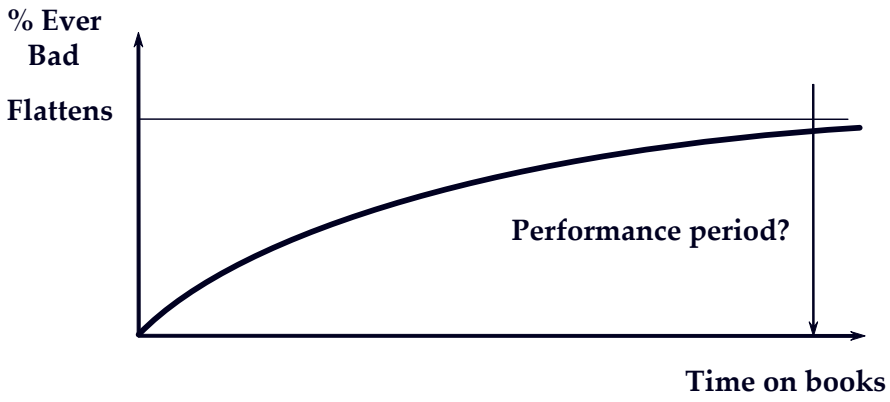

Figure 5.1 Selecting a Performance period

The solution to the Bad definition problem is therefore to find a proxy for write-off. The proxy must be reliable, stable, measurable and provide sufficient data. Bad is usually defined as a level of arrears; either current arrears or maximum arrears.

There are three ways this is often set:

- Historical precedence
- Management or consultant preferences
- Statistically.

The statistical approach is to use roll rates. Ideally these will be over the Performance period, but an annual probability of write-off will be better than nothing if the full analysis is not available.

Roll rates

Table 5.1 illustrates a range of potential Bad definitions for a scorecard development. The '% Write-off' column is the percentage of the cases that were in the category at the observation point that were written off 12 months later.

Definition	% Write-off
Currently 1 down	12%
Ever 1 down	8%
Currently2 down	33%
Ever 2 down	18%
Currently 3 down	56%
Ever 3 down	52%
Ever 2 x 2 down	20%
Ever 3 x 2 down	54%
Currently 2 down and ever 3+	54%
Currently 2 down and 3 x 1down	50%

Table 5.1 Roll rate analysis example

From this we see a number of potential Bad definitions such as 'currently 3 down' or 'ever 3 x 2 down' for example.

We could stop there, however the alternatives suggest that a combination of definitions might be appropriate. If we look at the definitions with over 50% and both rank them and make them hierarchical so that later definitions exclude accounts included in earlier ones we get the results in table 5.2.

Now we see that the 'currently 3 down' definition includes most of the cases that are picked up by the other definitions. However we can extend the definition of Bad to be 'currently 3 down or currently 2 down and ever 3'.

Definition	% Write-off
Currently 3 down	56%
Currently 2 down and ever 3	50%
Currently 2 down and 3 x 1 down	47%
Ever 3 down	31%
Ever 2 down	18%

Table 5.2 Hierarchical definition example

The practicality of the definition is not important. The business may track a totally different measure as a key quality metric. This doesn't matter. What does matter is that the bad rate the developer defines should be trackable. Scorecard performance – quite distinct from portfolio monitoring – should be based on the bad definition used to build the scorecard.

Intermediates

Intermediates are the 'grey' cases: the performance that is defined as neither Good nor Bad. Some developers define Intermediates and others refuse to have them.

The larger the proportion of Intermediates the more distinct the definitions of Good and Bad will be. The more distinct the performance definition, the higher the better the discrimination between Goods and Bads will be. However, a higher Gini does not necessarily mean a better scorecard. This is because the business will have Intermediates and will not have assessed these in trade-off considerations.

The advice is that if the business and developer are convinced that there should be an Intermediate category, then their impact should be immaterial. Keep Intermediates to less than 5% of the sample. In assessing the performance of the scorecard, the Intermediates should be added back in to the population. Look at the score distribution: where do the Intermediates lie. They will impact the acceptance rates so the 'run books' should be produced including the missing applications.

Exclusions
Not all applications will be considered in the development of a scorecard. Applications dropped are known as 'exclusions'.

The developer is trying to build a model to reflect a business decision. Certain cases will be inappropriate for inclusion for this assessment. Staff or VIPs are common examples.

A scorecard developed for assessment of 'premier' customers should not include 'standard' applications. If students do not get assessed by the scorecard, they should be excluded from the sample.

Other exclusions may be ones declined for policy reasons. For example if applicants with adverse credit are policy declined and will continue to be declined, then they should be excluded.

Fraudulent applicants are a category that often gets grouped with the Bads. However, the predictive attributes of frauds are different to other Bads and due to manipulation of application details the frauds can often look like Goods. It is best practice to identify and exclude frauds and develop a separate detection program for these applications.

Store cards suffer from a phenomenon known as 'hit and run'. This is where an applicant accepts the credit to obtain a discount, settles the account in full and never uses the facility again. IKANO Financial Services in the UK always excludes the 'hit and runs' from developments. They argue that identifying these customers as Good would encourage approval of such cases. Classifying them as Bad would undermine the scorecard since they look like the Goods. This works since IKANO track Bad rates that exclude the inactives. After the development they produce the 'run books' including and excluding the 'hit and runs' so that they understand the expected acceptance rates during discount and non-discount periods.

Conclusion
The determination of the definitions of Good and Bad and which cases should be included in the development is critical to the success of the scorecard. The starting point should always be: what are we trying to achieve and which customers represent the cases we want to distinguish between?

If in doubt, look at the score distribution of excluded categories and ask whether the scorecard decisions will be appropriate for the business objective.

The developer should document the choices and note the proportions in the original population and the final modelled sample. It is best practice to reproduce the final statistics including Intermediates and exclusions that reflect the population the

operation will see through-the-door. This will enable the business to make decisions that correspond to reality rather than the sample modelled.