



Profiling Pooled Bureau Data Samples

Credit Scoring and Credit Control XIII

C. Coleman
August 2013



Agenda

- Lender Challenge
- Pooled data
- Profiling Methodologies
- Weighting Models
- Conclusion





The Challenge



- Conservative Lenders have restricted/stopped lending due to economic conditions
- Some lenders looking to grow their books or launch new products to capitalise
- Limited historical account volumes means lenders unable to develop own models
- Pooled Bureau data can be used for model development
- Profiling may be necessary: adjusting one data file to look like another (not only volumes)



Pooled Samples

Pooled sample solutions can provide a great advantage to lenders with portfolios deemed unsuitable for a scorecard development based on their own data.

Portfolios may not be suitable for scorecard development due to

- Low volume of Bads
- Low total volumes
- New/Expanding product line
- Historically unstable portfolio

Creating a pooled sample profiled to the portfolio can be an effective solution

Sampled by:

- Portfolio & Date
- Similar Lenders*
- Volumes of Bad accounts

***Compliant Sample (No lender to contribute more than 25%)**





Pooled Samples Limitations

Though a Pooled Sample ensures sufficient accounts to develop a robust scorecard, there are limitations to this data source.

- Matching applications to opened accounts
- Unable to distinguish between Rejected & NTU
- Available fields/Segments
- Differing debt/collections strategies
- Varying live decline rates and cut-offs
- Target customers and marketing
- Application profile may differ

Sample profiling can be used to match a lender's "through the door" population using the available data.

Ensures developed model is appropriate for implementation and volumes are in line with expected live data – useful for setting cut offs

May not capture the behavioural profile of the target customers.



Sample Profiling

Profiling Options

There are several options available to manipulate the profile of a sample

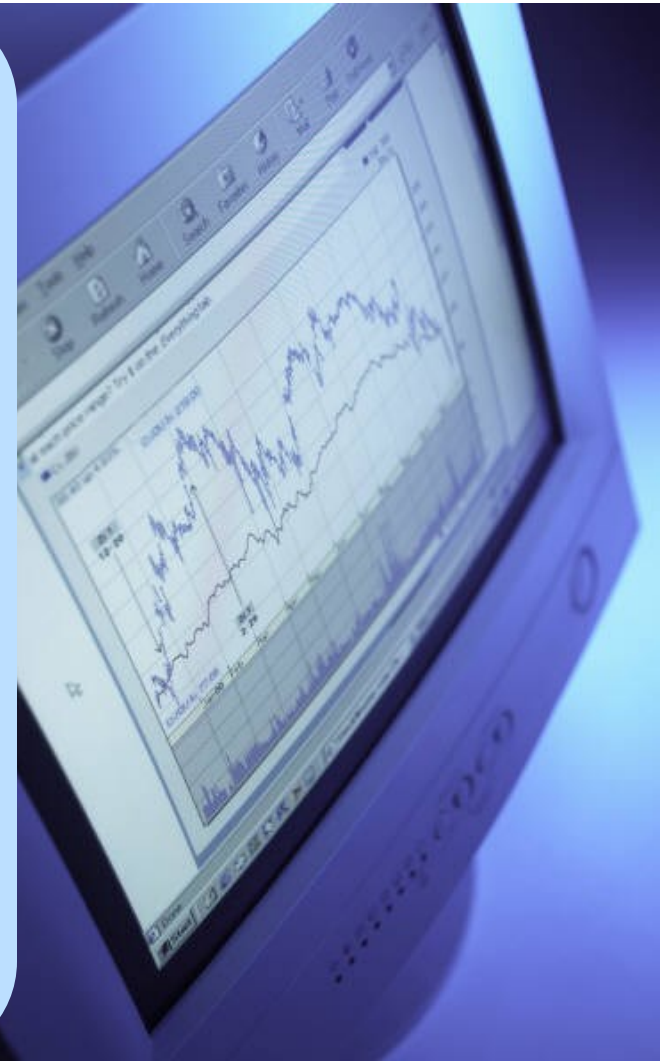
Method	Pros	Cons
Banded Variables: Re-weighting based on population proportion within each band	<ul style="list-style-type: none">• Simpler Method• Can be based on expected volumes (where actuals not available)	<ul style="list-style-type: none">• Results may be too simplistic• Banding of numeric values can lose accuracy• Difficult to incorporate multiple variables
Weighting Functions: Fitting a distribution function (e.g. Normal curve) to variable trend and creating a transformation	<ul style="list-style-type: none">• Simpler Method• Incorporates full trend of profiled variable	<ul style="list-style-type: none">• Results may be too simplistic• Difficult to incorporate multiple variables
Weighting Models: Regression model to predict how similar the lender profile is to the pooled data	<ul style="list-style-type: none">• incorporates multiple variables well• Better profile alignment achievable	<ul style="list-style-type: none">• Complex method• Can alter the performance trends• Time consuming



Sample Profiling Logistic Model

Logistic Model Profiling:

- A model used to identify the pooled applicants most similar to the bespoke lender
- Requires sample of lender application data
- A binary outcome is defined to determine whether a record is bespoke lender application (outcome = 1) or a pooled application (outcome = 0)
- Top performing variables considered for the modelling
- Gini coefficient measures how dissimilar the pooled data is from the lender specific sample prior to the profile weighting



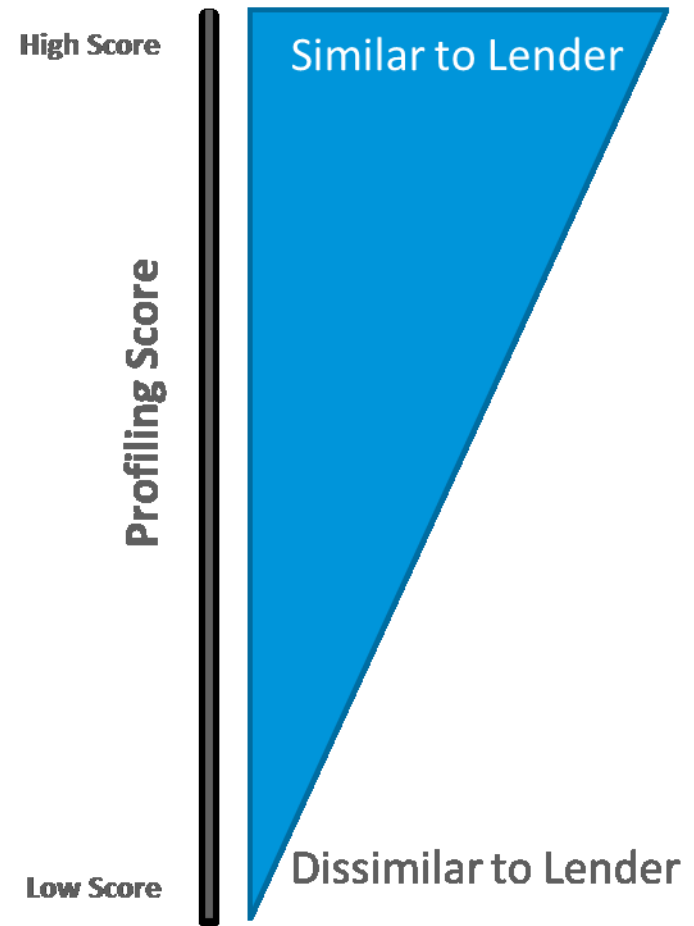


Sample Profiling Logistic Model

- Pooled data with a high model score is more likely to be 'similar' to lender data and weighted up.
- Conversely, a low score shows the observation is easily discriminated from lender data and hence has a different profile and should have its influence reduced.
- The raw score from the logistic model can be easily converted into a probability using the standard logit transformation:

$$p(\text{lender}) = \frac{1}{1 + e^{-x}}$$

- This value is used as part of the new weight.





Sample Profiling Re-weighting Algorithm

- Once the profiling model has been built, its output probability is used within a re-weighting function and applied to the pooled sample
- Using a normalised weighting factor ensures that the overall volume of the generic sample is maintained.

Normalised Weighting factor:

$$\text{Weight}_{\text{New}} = \text{Weight}_{\text{Old}} \times \frac{P(\text{lender})}{\text{Average}.P(\text{lender})_{\text{PooledSample}}}$$

Alternative Weighting factors can be used

- Near-Normalised Weighting Factor
- Non-Normalised Weighting Factor





Sample Profiling

Iterating the Algorithm

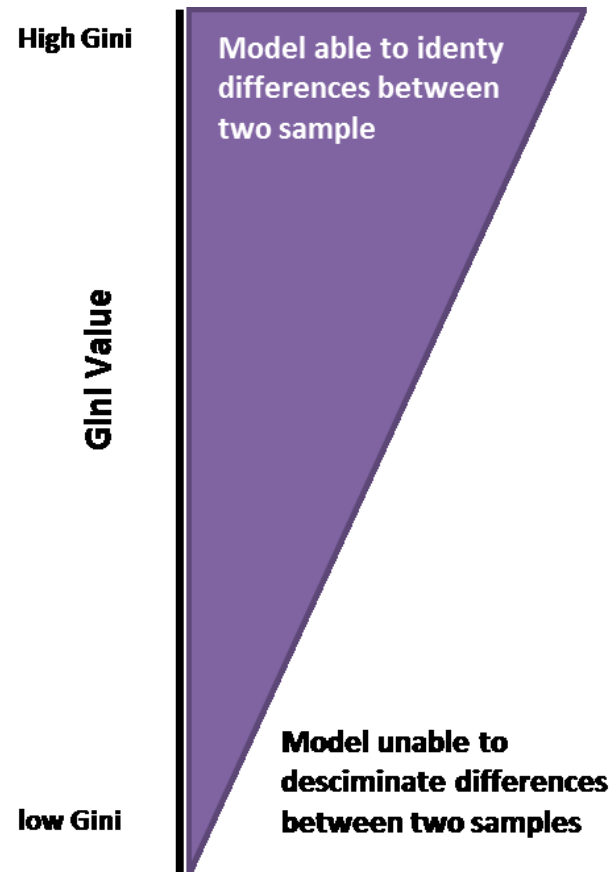
- Although the algorithm will in most cases converge quickly, it may be necessary to undertake a number of iterations of this process on the new weight.
- The table below gives an example of results.

Iteration Number	Gini of Model
1	60
2	20
3	10
4	5

- Each iteration of the model indicates less discrimination between the Pooled and the Lender Sample

Conclusion:

Low Gini → Profile of Pooled sample ≈ Profile of Lender Sample





Sample Profiling

Profiling Model: Tests, Validations and Limitation



Tests and Validations

- Score Distribution and Characteristic Analysis reports across key scores and variables before and after the profiling gives insight in to the performance trends of the sample as a result of the profiling
- Expect to see improvement in the volume alignment of non-model variables
- Check for:
 - Extreme weights on Goods and Bads
 - Inverse or artificial trends on Key Variables
 - Differences in how variables are populated across the two samples

Limitations

- The observed bad rate could be reduced or increased by applying the new weighting



Conclusions

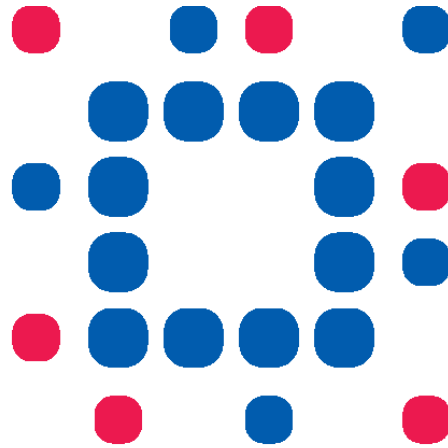
Pooled data samples can provide a stable basis for the development of bespoke scoring models

Profiling may not significantly improve the models ability to rank risk

Sample profiling is recommended where alignment is key e.g. setting cut offs based on previous sample volumes, forecasting bad debt and can give greater insight of live results.

The use of models can greatly improve the alignment of a sample profile over other weighting methods although they can introduce other issues.





Experian™

A world of insight