A misleading extrapolation of reject performance from the accepts

Ross Gayler

31 August 2007

CSCC X

Summary

- Reject inference: How would this application have performed if it had been accepted?
- Two broad classes of approach to reject inference:
  1. Outcome information only for the accepts
  2. Alternative (proxy) outcome information for all applications
- Approach 1 is essentially an extrapolation from the accepts
- Approach 1 does not allow assessment of the extrapolation

  (Because there is no outcome information for the rejects)
- This is a case study comparing the two approaches
- This study clearly illustrates the risk in blind extrapolation

I assume that the audience is broadly familiar with reject inference.

The goal of RI is to estimate how the rejects would have performed if they had been accepted.

Approaches to RI may be categorised by whether they use a proxy outcome that can be observed for accepts and rejects.

Approaches without a proxy outcome are essentially extrapolation from behaviour observed on the accepts only. (Observe the performance pattern in the accepts and extrapolate it to the rejects – possibly with a subjective modification.)

The trouble with extrapolation approaches is that there is no way to check the accuracy of the extrapolation.

This case study contrasts the two approaches to reject inference in one project.

This particular data set shows that it is possible to be quite misled by extrapolating from the accepts.

Structure of presentation

veda ADVANTAGE

- Data set
- Bureau outcome as a proxy for Lender outcome
- High-side overrides
- Reject inference
- Alternative models

Outline of the presentation:

Describe the project and data

Show that bureau outcome can be used as a proxy for lender outcome.

Describe the high-side overrides as they are very important to the argument here.

Contrast reject inference performed both ways.

Explore the consequences for the lender of the alternative reject inferences.

| Data set

Data set

- The same applications used by the Lender to develop a model
  - Consumer lending product, Point of sale data
- Veda were given a restricted subset of variables
  - Old application score
  - New application score
  - Outcome (Lender's Good/Bad flag for accepts)
  - Lender's inferred reject performance (probability)
  - Identity (for retrospective bureau matching)
- 80/20 Development / Validation split used by the Lender

This case study is based on model development performed by a lender.

The lender used an extrapolation RI method. (The lender did not have access to bureau outcome data for rejects).

The Veda Advantage project aimed to demonstrate the value of bureau outcome data for reject inference.

The lender supplied a subset of their development data set: all observations but only a restricted set of variables.

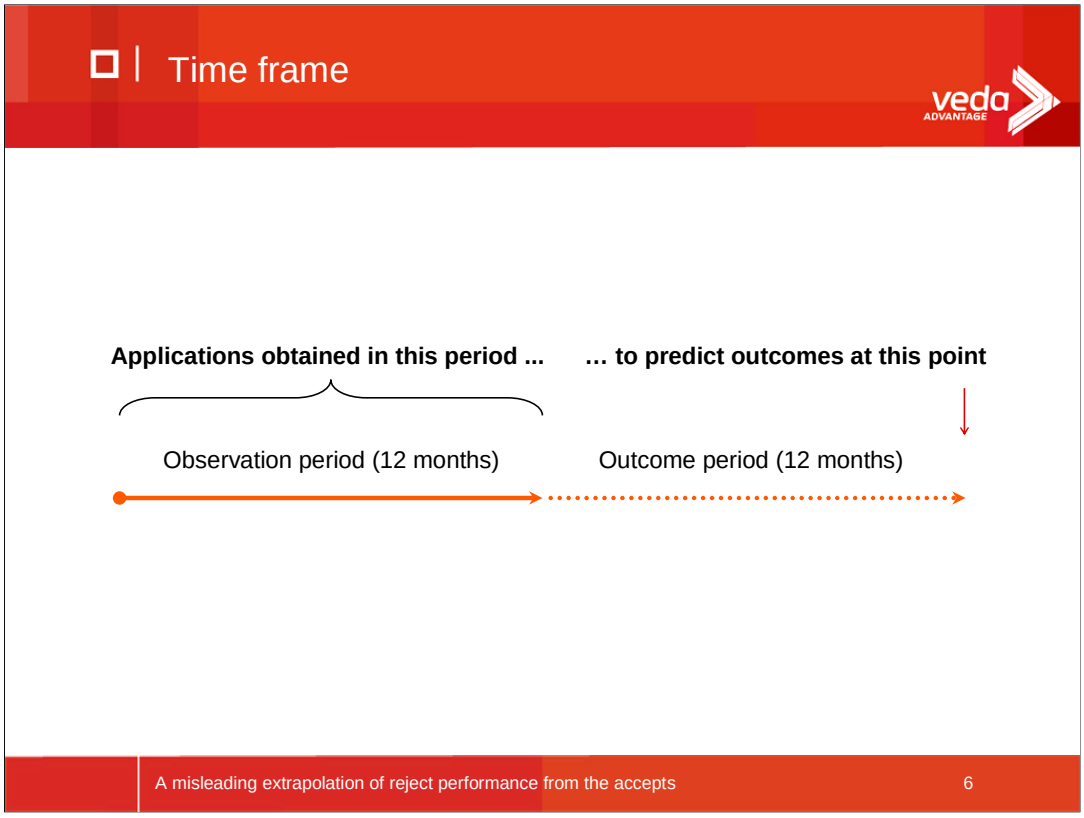
-Old and New scores (but not scorecards)

-Lender's actual outcome for accepts

-Lender's inferred outcome for rejects (as a probability)

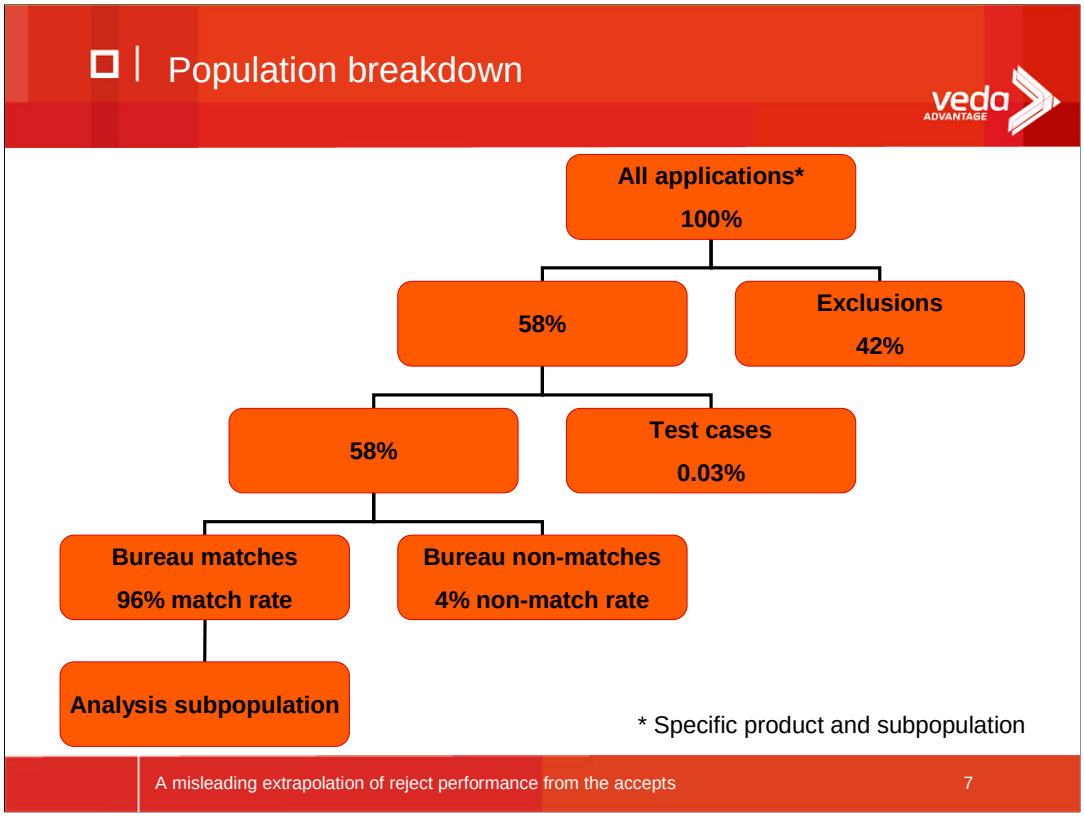
-Identity for bureau matching

-Development/validation split



Observations were all applications for some product received over 12 months.

Outcome was measured at a fixed point 12 months after the end of the application window. (The individual outcome periods were 12 to 24 months.)

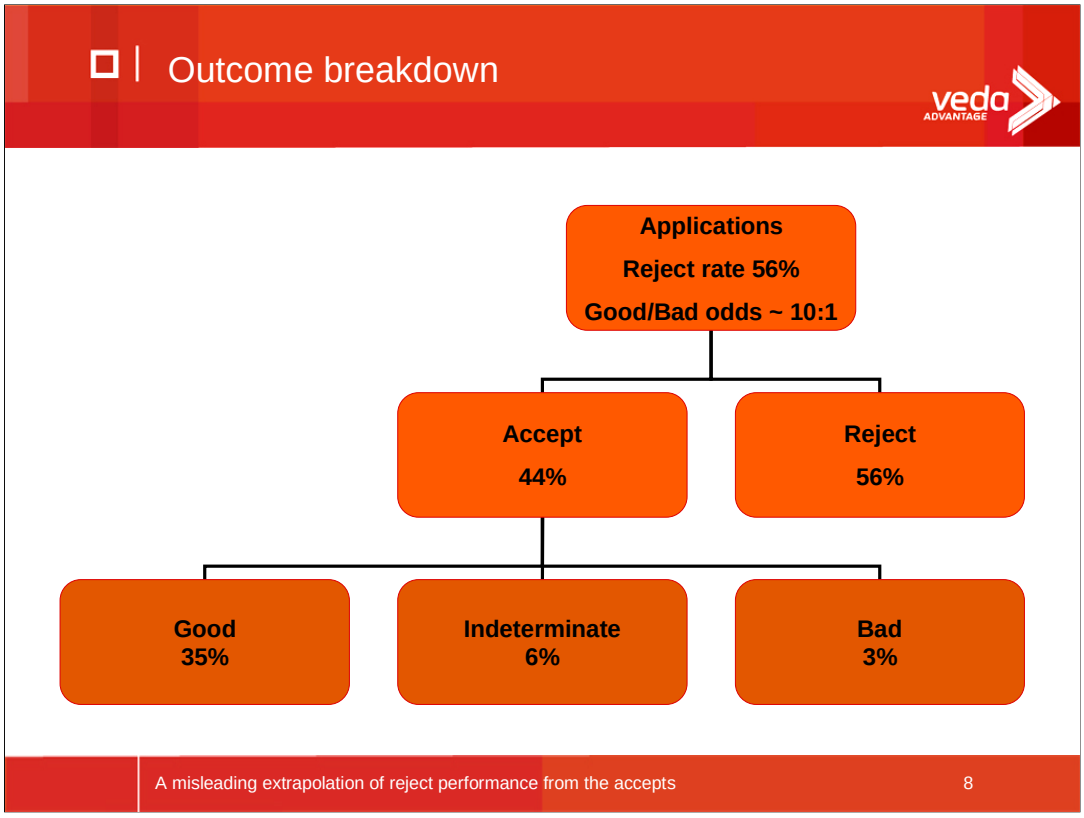


The development sample records came from a specific subpopulation of the applications for a specific credit product.

42% of applications were declined because of policy rules that would be retained after the scorecard redevelopment.

A small number of test cases (e.g. same address as the lender's head office, names like "Joe Test") were identified and excluded from the analysis. It appears that these cases were included in the lender's development sample. The detection of these cases is not really material to the outcome but does illustrate the types of data quality issues that arise.

Applications were matched to bureau records to obtain bureau outcome information. The match rate was quite high. We assume that the matched applications are representative of the non-matched applications with respect to the ensuing RI. This would be more of an issue if the match rate were lower.



The reject rate is quite high (56%). Therefore, RI is important.

The Good/Bad definition was based on delinquency in the outcome period. The account did not need to default or write off in order to be classed as Bad.

Indeterminate is intermediate between Good and Bad.

Bureau outcome as a proxy for Lender outcome

veda ADVANTAGE

Lender & Bureau outcomes (All applications)

Lender outcome	Bureau Bad rate
G	5%
I	12%
R	26%
B	74%

Lender outcome	Bureau Good (Col. %)	Bureau Bad (Col. %)
G	42%	10%
I	6%	4%
R	50%	74%
B	1%	13%

A misleading extrapolation of reject performance from the accepts 10

We are going to use bureau outcome as a proxy for lender outcome, so we need to demonstrate that they are related.

The left hand table shows the observed bad rate (defaults) on the bureau as a function of the lender outcome category. The right hand table shows the distribution across lender outcome categories within each of the bureau outcome categories. These show the expected patterns.

Lender & Bureau outcomes (Accepts only)

Row %	Bureau Good	Bureau Bad
Lender Good*	94%	6%
Lender Bad	26%	74%

Col. %	Bureau Good	Bureau Bad
Lender Good	98%	51%
Lender Bad	2%	49%

* Lender Good includes Indeterminate for subsequent analyses

Lender & Bureau outcomes (Accepts only)

Using logistic regression to predict one outcome from the other

Regression predictor	Regression outcome	Odds ratio	Gini
Bureau outcome	Lender outcome	41	0.67
Lender outcome	Bureau outcome	41	0.47

Lender & Bureau outcomes (Accepts only)

Predictability of the Lender outcome from the Bureau outcome and the available scores (at the point of application)

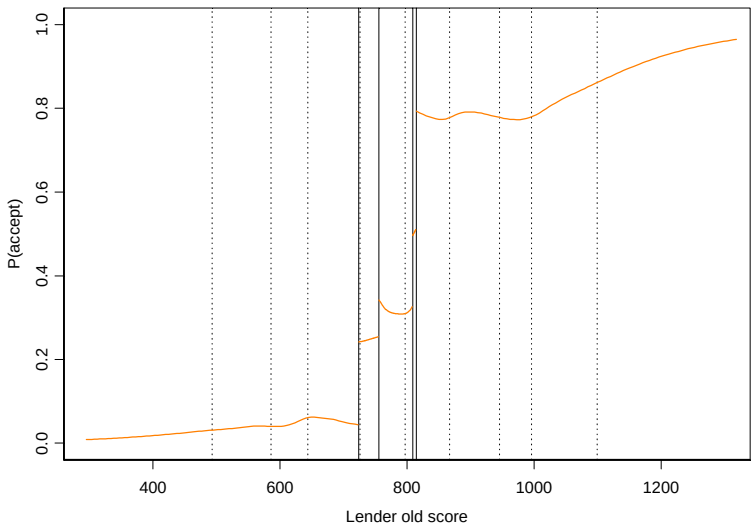
Predictor	Bureau outcome	Lender new score (total)	Lender old score	Bureau score (Consumer)	Lender new score (bureau component)	Bureau score (Value Add)
Gini	0.67	0.40	0.30	0.29	0.22	0.19

Bureau outcome as a proxy for Lender outcome

- Lender and Bureau outcomes are related (as expected)
- Bureau outcome is a better predictor of Lender than vice versa
  - Bureau outcome contains extra information (other defaults)
- Bureau outcome is a better predictor of Lender outcome than any available score
- Bureau outcome was not available to the Lender for reject inference

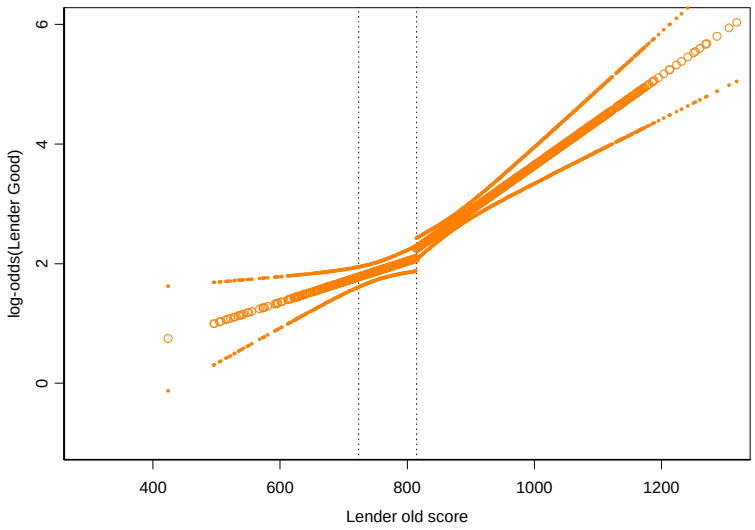
| High-side overrides

Accept rate as a function of old score



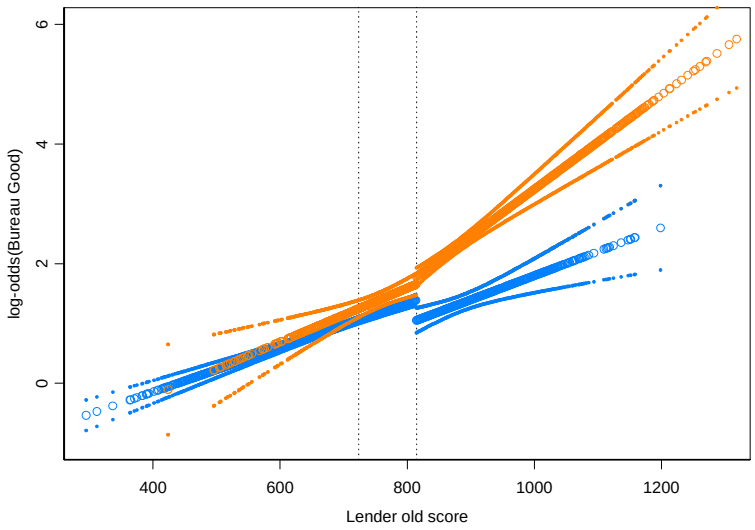
A misleading extrapolation of reject performance from the accepts

Lender outcome as a function of old score (accepts only)



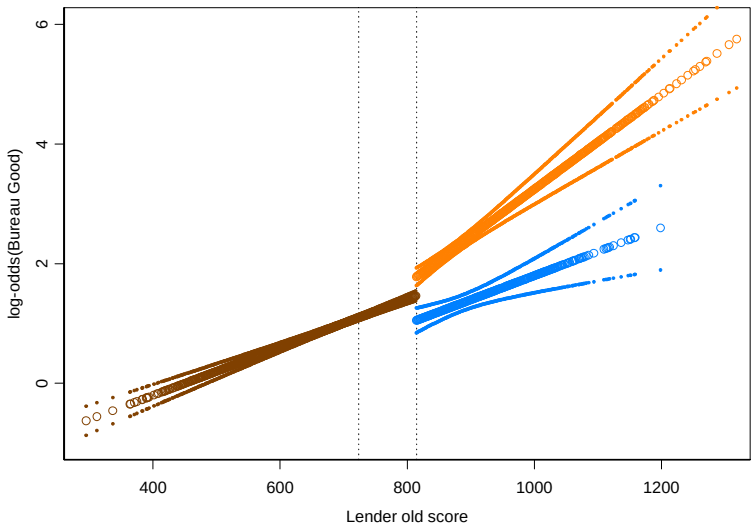
A misleading extrapolation of reject performance from the accepts

Bureau outcome as a function of old score (accepts & rejects)



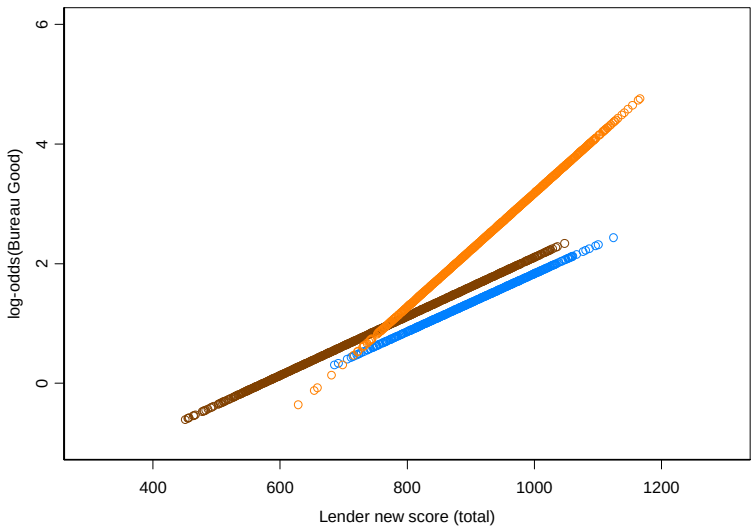
A misleading extrapolation of reject performance from the accepts

Bureau outcome as a function of old score (accepts & rejects)



A misleading extrapolation of reject performance from the accepts

Bureau outcome as a function of new score (accepts & rejects)



A misleading extrapolation of reject performance from the accepts

Conclusions from calibrations

- Calibration of Lender outcome to old score shows clear evidence of selection bias
- Calibration of Bureau outcome parallels calibration of Lender outcome
- Low-side overriding independent of risk
- Low side calibrations show granting credit card does not increase bureau risk
- High-side overriding is sensitive to risk
- High-side override predictive information is not incorporated in Lender or Bureau scores

| Reject inference

Reject inference

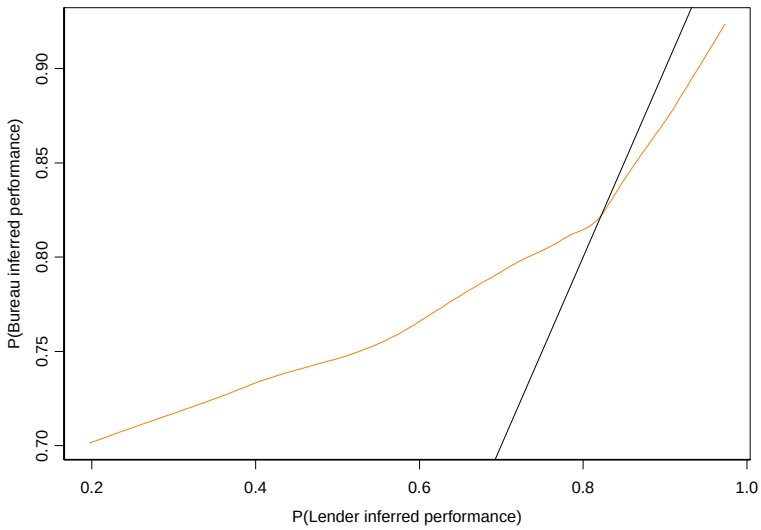
- Develop model to predict Lender outcome from Bureau outcome
- Predictive model is based only on the accepts

  (because they have Lender and Bureau outcome)
- Apply the model to the rejects

  Use Bureau outcome to predict probabilistic Lender outcome for rejects

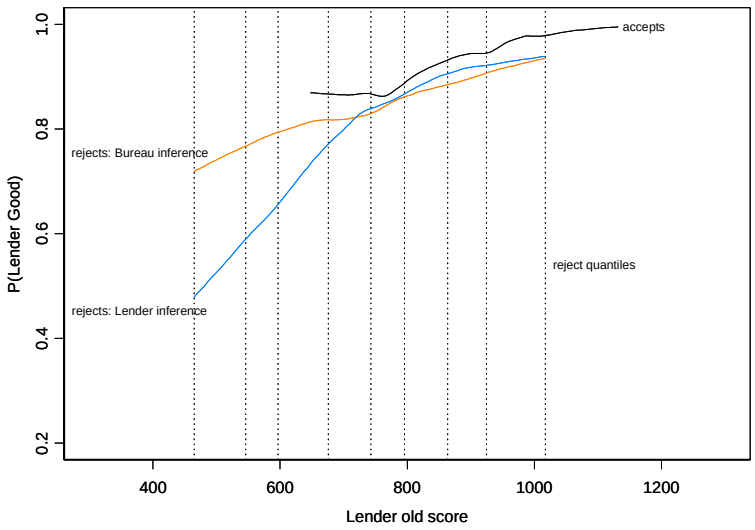
  Randomly realise the inferred Lender outcome for the rejects

Comparison of inferred outcomes (rejects)



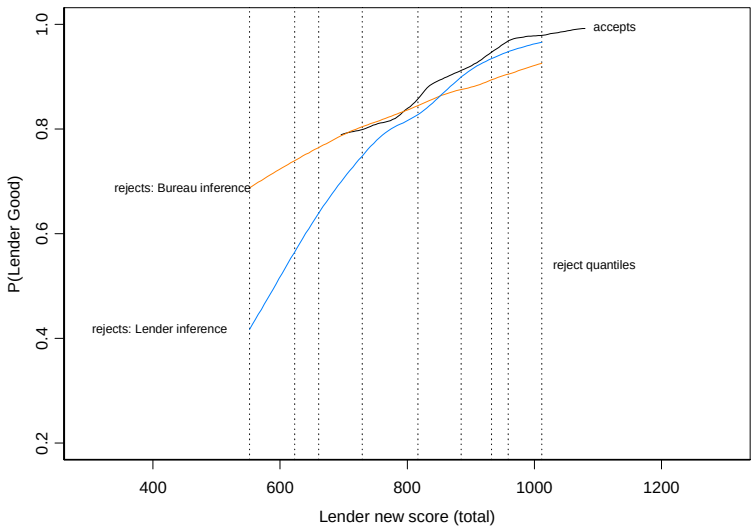
A misleading extrapolation of reject performance from the accepts

Comparison of actual and inferred outcomes



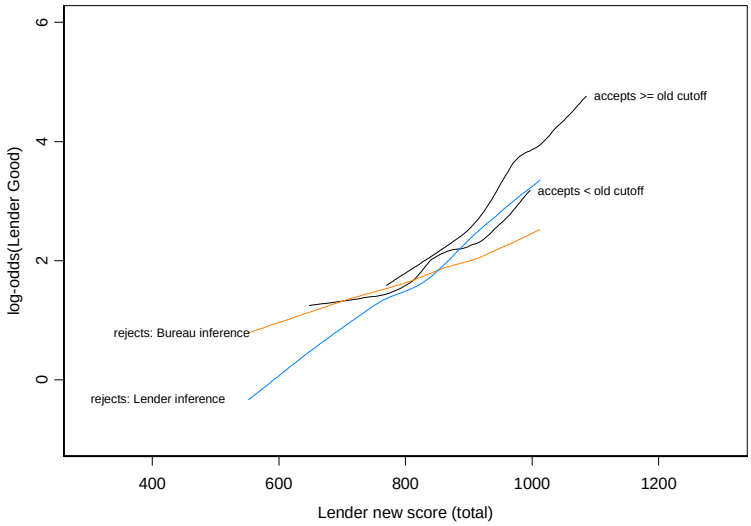
A misleading extrapolation of reject performance from the accepts

Comparison of actual and inferred outcomes



A misleading extrapolation of reject performance from the accepts

Comparison of actual and inferred outcomes



A misleading extrapolation of reject performance from the accepts

Conclusions from reject inference comparison

- Lender range of inferred performance is too stretched
  - Lender inference shows greater outcome range than Bureau
  - Lender inference based on weaker predictors
  - Bureau inference based on real reject outcomes
- Lender rejects > old cutoff are more similar to accepts
  - Contrary to Bureau outcome
- Lender rejects < old cutoff are much worse than accepts
  - Contrary to Bureau outcome
- Over-optimistic swap-set
  - Makes new Lender model look better than it is

| Alternative models

Construction of new models

- Use known Lender outcome for accepts
- Use Bureau inferred outcome for rejects
- Fit logistic regression models from available scores to known plus inferred outcome for accepts and rejects
- Exclude Indeterminates from regression (as the Lender did)
- Fit models on development sample
- Assess models on validation sample

Comparison of models (Gini* on accepts + rejects)

Model	Lender outcome	Bureau outcome
Lender old	0.359	0.269
Lender new	0.458	0.337
Lender new + Bureau value-add	0.457	0.362
Lender new + Bureau consumer	0.464	0.390

* Results from Validation set only

Comparison of models
($Gini_i$ / $Gini_j$* on accepts + rejects)

Model	Lender outcome	Bureau outcome
Lender old	0.78	0.80
Lender new	1.00	1.00
Lender new + Bureau value-add	1.00	1.07
Lender new + Bureau consumer	1.01	1.16

* Results from Validation set only

Comparison of models (Gini* on accepts only)

Model	Lender outcome	Bureau outcome
Lender old	0.281	
Lender new	0.370	
Lender new + Bureau value-add	0.375	0.370
Lender new + Bureau consumer	0.405	0.433

* Results from Validation set only

Comparison of models ($Gini_i/Gini_j$* on accepts only)

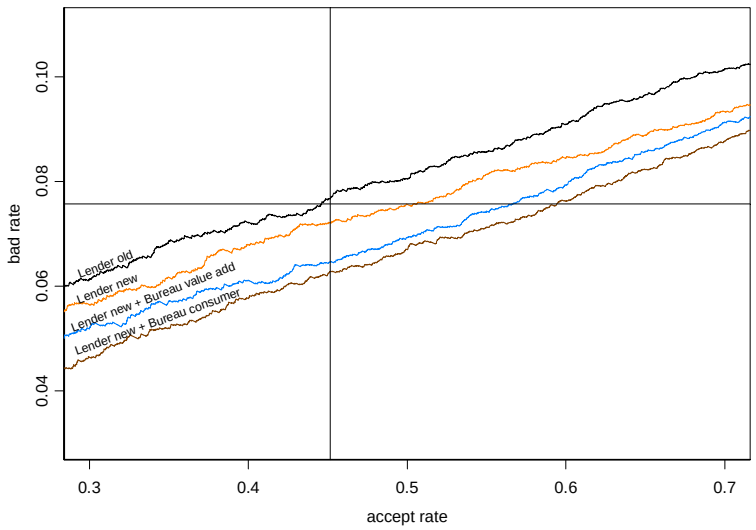
Model	Lender outcome	Bureau outcome
Lender old	0.76	
Lender new	1.00	
Lender new + Bureau value-add	1.01	1.00
Lender new + Bureau consumer	1.09	1.17

* Results from Validation set only

Conclusions from comparison of model Ginis

- Lender outcome overstates predictive power
  - Ginis generally higher
  - Lender inferred outcome range is stretched relative to Bureau inferred outcome range
- Bureau scores do add power for through-the-door population
  - For accepts only when assessed against Lender outcome
  - For all applications when assessed against Bureau outcome

Accept rate vs Bad rate tradeoffs



A misleading extrapolation of reject performance from the accepts

Total profit over 3 years (same accept rate)

Model	%
Lender old	83
Lender old + overrides	87
Lender new (total)	100
Lender new + Bureau value add	127
Lender new + Bureau consumer	134