Best Practices in Reject Inferencing

Dennis Ash
Experian

Steve Meester
The CIT Group

Conference on Credit Risk Modeling and Decisioning
Wharton FIC, University of Pennsylvania
May 29-30, 2002
Objectives

Introduction

- What is Reject Inference
- Why we Need Reject Inference
- Literature Review

Reject Inference Techniques

- Description
- Assumptions
- Outcomes
Reject inference: What is it?

Assignment of an inferred status (G/B) to applicants declined for credit.

Equivalent to saying “if these applicants had been accepted, this is how they would have performed”
New applicant scoring

Predict outcome status

Application dates

June ‘00

Approved

Declined

May ‘01

Outcome point

Good

Bad

May ‘02
Is the missing outcome performance for rejects a problem?

- Sample bias
- Need statistically sound representative scorecard development sample
- Need scorecard to be effective for applications with reject profile
- Depends on past decision making
Why we need it

- If prior screening process used by the lending institution to separate applicants into accepts and rejects was applied in a (stratified) random manner.
- Then all applicants would be represented in the accepted population.
Why we need it

- A good (stratified) random sample of accepts could then represent the applicant pool.
- It would contain some occurrences of bad credit followed by bad performance for all regions of the applicant pool.
Why we need it

- Then we would not need reject inferencing.
- This is not often done. It is too expensive because the losses are too high.
Literature Review

Literature Review

Literature Review


Why do we need reject inferencing?

- Development sample bias
- Forecast bias
Reject inference techniques
Techniques

- No reject inference
- Re-classification
- Re-weighting
- Parcelling
- Heckman’s bias correction
- Supplemental Bureau Data
No reject inference

- Build model on known bad / good flag
- Ignore rejects in model development
- Incorporate rejects in forecast
Reclassification

- The worst cases of rejects are selected and reclassified as accepts
- A “bad” status is then assigned
Reclassification – How’s

The rejects are selected by

- Reject / accept model
- Serious derogatory information
Reclassification – How’s

Reject / accept model

- Used to identify the worst rejected applicants
- Apply reject / accept model to approved and rejected accounts
- The lowest scoring rejects are reclassified
Reclassification – How’s

Serious derogatory information

- Used to identify the worst rejected applicants
- Rejects who have more than a significant number of trades with seriously derogatory information
- Analyze RA and BG cross-tabs
Re-weighting

- Based on accept extrapolation
  - Accepted accounts are similar to declines
  - How declines would have performed if approved
- Accepts are weighted up to represent the rejects
Re-weighting – How’s

Reject / accept model

- Used to identify similar applicants
- Apply reject / accept model to approved and rejected accounts
- The accounts (rejected and approved) are grouped by similar score
- The behavior of the approved accounts in a score interval can be used to infer what the likely behavior of the corresponding rejects would be, had they been approved
## Re-weighting – Example

<table>
<thead>
<tr>
<th>Score Interval</th>
<th>Rejects</th>
<th>Accepts=</th>
<th>Bads+</th>
<th>Goods</th>
</tr>
</thead>
<tbody>
<tr>
<td>601-700</td>
<td>20</td>
<td>100</td>
<td>10</td>
<td>90</td>
</tr>
</tbody>
</table>
Re-weighting – Example

- 90% of approved accounts were good, while 10% were bad
- Can infer that
  - 10% of rejects in that interval (0.10*20=2) might have gone bad, had they been approved
  - 90% of reject (0.90*20=18) would be good
- By weighting the approved accounts by 1.2 (120/100) the sample would contain
  - 12 bads and 108 goods
- Therefore, the approved accounts were used as proxies for the rejects
Parceling

- Rejects are assigned into both bad and good categories, or probability of good
- Based on logical and statistical evidence of the proportion that would have gone bad
Parceling illustration

- Build reject / accept model
- Build known good / bad model
- Plot known good / bad model versus reject / accept model
  - Accepts
  - Rejects
- Adjust performance for rejects to reflect trend
GB model based on known goods and bads

![Graph showing PGB and PAR values]
GB model applied to rejects
Adjusted performance on rejects
Heckman’s Correction - Introduction

Hand & Henley (1993)

- Lack of theoretical foundation that could justify any claim of bias correction
- Additional assumptions could validate RI methods, only if they are reasonable and consistent with statistical theory
Heckman’s correction

Heckman (1979)

- Discussed bias from using nonrandom selected samples to estimate behavioral relationships as a specification error
- He suggests a two stage estimation method to correct the bias
- The correction is easy to implement and has a firm basis in statistical theory
Heckman’s correction

- Normality assumption
- Provides a test for sample selection bias
- Formula for bias corrected model
Shortcomings/Assumptions
No Reject Inference

- Does not adjust for sample bias.
- Portfolio quality estimates will be optimistic over the rejects.
Reclassification

- Ad-hoc.
- Implies \( P(\text{bad} \mid X) = 1 \) over a segment of the covariate space. We know this is not true.
- May bias the scoring model over the accepts.
Re-weighting

- Assumes

\[ P(\text{bad} \mid X, \text{rejected}) = P(\text{bad} \mid X, \text{accepted}). \]

This is a very strong and generally unrealistic assumption.

Implies accept/reject procedure provides no discrimination given the bureau data \( X \).

- Must have accepts with the same bureau profile as the rejects.
Heckman/Bivariate Probit

- Accept/reject procedure must be stochastic.

- All factors used in the accept/reject decision must be observable, i.e. no additional factors may be considered by credit managers.
How Well Do These Work?

- Several studies have shown that gains from using correction for sample selection based on observation data are less than expected.
- Reliable model based reject inference is impossible - model assumptions are important and are violated.
- But the information loss due to selection bias is substantial.
- Need real information on rejects.
Supplemental Bureau Data

- Obtain bureau data on accepts and rejects at the end of the observation period.

- Use the performance with other creditors over the observation period to infer how the rejects would have performed had they been accepted.
Supplemental Bureau Data Methodology

- Let $Z$ denote the downstream bureau data.
- Fit a model for $P(\text{bad} \mid Z)$ over the accepts.
- Impute $P(\text{bad} \mid Z)$ for the unobserved $Y$ for a reject.
- This is parceling BUT we use payment performance with other creditors over the time frame of interest to determine the parceling for a prior decline.
- The parceling is no longer subjective. It is driven by supplemental performance data.
Assumptions

**Key assumption:**

\[ P(\text{bad} \mid X, Z, \text{rejected}) = P(\text{bad} \mid X, Z). \]

That is, the bureau data at time of application and the downstream bureau data contain all the relevant information about \( P(\text{bad}) \).

This is a much weaker assumption than required for re-weighting.
Shortcomings

Requires a good bureau match rate.
Supplemental Bureau Data: Cautions

- Models for which the likelihood score is linear in Y, just impute $P(\text{bad} \mid X, Z)$ for Y.
  - e.g. logistic regression model.

- Models for which the likelihood score is non-linear in Y, impute $E[S(\theta) \mid X, Z]$ for $S(\theta)$, Meester (2002).
  - e.g. linear model.

- Naive standard error estimates are not correct. Bootstrap!
Example

- 9259 leases from a business which approves approx. 98% of applications.

- Create “declines” if any prior liens or judgments.

<table>
<thead>
<tr>
<th>“Accepts”</th>
<th>“Declines”</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td># apps.</td>
<td>8,127</td>
<td>1,132</td>
</tr>
<tr>
<td>Bad rate</td>
<td>6.3%</td>
<td>13%</td>
</tr>
</tbody>
</table>
Example

- Fit logistic regression model to full sample with observed response to get the “Gold Standard” model.

- Fit model with no reject inference.

- Fit logistic regression model using the reject inference procedure.
Example: Parameter Estimates

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.250</td>
<td>-2.2161</td>
<td>-2.1472</td>
</tr>
<tr>
<td>log(liens+1)</td>
<td>0.7378</td>
<td>NA</td>
<td>0.7824</td>
</tr>
<tr>
<td>1(judgments&gt;0)</td>
<td>0.7143</td>
<td>NA</td>
<td>1.0063</td>
</tr>
<tr>
<td>X3</td>
<td>0.4668</td>
<td>0.4497</td>
<td>0.4299</td>
</tr>
<tr>
<td>X4</td>
<td>-0.2303</td>
<td>-0.1911</td>
<td>-0.1919</td>
</tr>
<tr>
<td>X5</td>
<td>-0.4052</td>
<td>-0.4744</td>
<td>-0.4100</td>
</tr>
<tr>
<td>X6</td>
<td>0.7744</td>
<td>0.9400</td>
<td>0.7634</td>
</tr>
<tr>
<td>X7_1</td>
<td>0.8031</td>
<td>0.7371</td>
<td>0.5883</td>
</tr>
<tr>
<td>X7_2</td>
<td>0.5916</td>
<td>0.4677</td>
<td>0.3423</td>
</tr>
<tr>
<td>X7_3</td>
<td>0.9795</td>
<td>1.1561</td>
<td>0.9778</td>
</tr>
<tr>
<td>X7_4</td>
<td>1.1023</td>
<td>1.0388</td>
<td>0.9592</td>
</tr>
<tr>
<td>log(suits+1)</td>
<td>NA</td>
<td>0.4953</td>
<td>NA</td>
</tr>
</tbody>
</table>
## Estimated Standard Errors

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Naive Estimate</th>
<th>Bootstrapped</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.3726</td>
<td>0.3638</td>
</tr>
<tr>
<td>log(liens+1)</td>
<td>0.0963</td>
<td>0.1688</td>
</tr>
<tr>
<td>1(judgments&gt;0)</td>
<td>0.1318</td>
<td>0.4030</td>
</tr>
<tr>
<td>X3</td>
<td>0.1910</td>
<td>0.1922</td>
</tr>
<tr>
<td>X4</td>
<td>0.0275</td>
<td>0.04164</td>
</tr>
<tr>
<td>X5</td>
<td>0.0445</td>
<td>0.0444</td>
</tr>
<tr>
<td>X6</td>
<td>0.1302</td>
<td>0.1217</td>
</tr>
<tr>
<td>X7_1</td>
<td>0.3667</td>
<td>0.3734</td>
</tr>
<tr>
<td>X7_2</td>
<td>0.3738</td>
<td>0.3875</td>
</tr>
<tr>
<td>X7_3</td>
<td>0.3871</td>
<td>0.3955</td>
</tr>
<tr>
<td>X7_4</td>
<td>0.3496</td>
<td>0.3639</td>
</tr>
</tbody>
</table>
## Portfolio Quality: Percent Bad

<table>
<thead>
<tr>
<th>Approval Rate</th>
<th>Full Data Model Actual</th>
<th>Imputed Model Actual</th>
<th>Estimate</th>
<th>No Reject Inference Actual</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>5.9</td>
<td>6.0</td>
<td>6.1</td>
<td>6.2</td>
<td>5.3</td>
</tr>
<tr>
<td>80%</td>
<td>5.0</td>
<td>5.1</td>
<td>5.2</td>
<td>5.3</td>
<td>4.5</td>
</tr>
<tr>
<td>70%</td>
<td>4.1</td>
<td>4.3</td>
<td>4.3</td>
<td>4.5</td>
<td>3.7</td>
</tr>
<tr>
<td>60%</td>
<td>3.6</td>
<td>3.6</td>
<td>3.6</td>
<td>3.8</td>
<td>3.2</td>
</tr>
<tr>
<td>50%</td>
<td>3.2</td>
<td>3.2</td>
<td>3.2</td>
<td>3.4</td>
<td>2.9</td>
</tr>
<tr>
<td>40%</td>
<td>2.7</td>
<td>2.6</td>
<td>2.7</td>
<td>3.0</td>
<td>2.4</td>
</tr>
<tr>
<td>30%</td>
<td>2.4</td>
<td>2.4</td>
<td>2.4</td>
<td>2.7</td>
<td>2.2</td>
</tr>
<tr>
<td>20%</td>
<td>2.3</td>
<td>2.1</td>
<td>2.1</td>
<td>2.8</td>
<td>2.2</td>
</tr>
<tr>
<td>10%</td>
<td>2.6</td>
<td>2.3</td>
<td>2.3</td>
<td>2.3</td>
<td>1.9</td>
</tr>
</tbody>
</table>
Conclusions

• Other reject inference methods
  ■ require very restrictive assumptions: Heckman/Bivariate Probit, Re-weighting;
  or
  ■ employ adhoc intervention which may lead one astray: Re-classification, parceling.
Conclusions

-Parceling with downstream bureau data uses additional, *data driven* information, for the reject inference.
-Requires fewer assumptions.
-Requires good bureau match.
### When to use each method

<table>
<thead>
<tr>
<th>Method</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>- Low reject rate</td>
</tr>
<tr>
<td></td>
<td>- Random decisioning</td>
</tr>
<tr>
<td></td>
<td>- Simulate policy rules</td>
</tr>
<tr>
<td></td>
<td>- Reclassification with derogs</td>
</tr>
<tr>
<td></td>
<td>- Reclassification with score</td>
</tr>
<tr>
<td></td>
<td>- Extreme rejects by score</td>
</tr>
<tr>
<td></td>
<td>- Re-weighting</td>
</tr>
<tr>
<td></td>
<td>- Accept / reject overlap</td>
</tr>
<tr>
<td></td>
<td>- Parceling</td>
</tr>
<tr>
<td></td>
<td>- Strict adherence to scorecard</td>
</tr>
<tr>
<td></td>
<td>- Heckman</td>
</tr>
<tr>
<td></td>
<td>- To test bias</td>
</tr>
<tr>
<td></td>
<td>- Easy to use</td>
</tr>
<tr>
<td></td>
<td>- Bureau</td>
</tr>
<tr>
<td></td>
<td>- Quality bureau match</td>
</tr>
</tbody>
</table>
## When to use each method

<table>
<thead>
<tr>
<th>Method</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>- Low reject rate</td>
</tr>
<tr>
<td></td>
<td>- Random decisioning</td>
</tr>
<tr>
<td>Reclassification with derogs</td>
<td>- Simulate policy rules</td>
</tr>
</tbody>
</table>
## When to use each method

<table>
<thead>
<tr>
<th>Method</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>■ Low reject rate</td>
</tr>
<tr>
<td></td>
<td>■ Random decisioning</td>
</tr>
<tr>
<td>Reclassification with derogs</td>
<td>■ Simulate policy rules</td>
</tr>
<tr>
<td>Reclassification with score</td>
<td>■ Extreme rejects by score</td>
</tr>
<tr>
<td>Parceling</td>
<td></td>
</tr>
<tr>
<td>Strict adherence to scorecard</td>
<td></td>
</tr>
<tr>
<td>Heckman</td>
<td></td>
</tr>
<tr>
<td>Accept / reject overlap</td>
<td></td>
</tr>
<tr>
<td>Bureau</td>
<td></td>
</tr>
<tr>
<td>Quality bureau match</td>
<td></td>
</tr>
</tbody>
</table>
### When to use each method

<table>
<thead>
<tr>
<th>Method</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>Low reject rate</td>
</tr>
<tr>
<td></td>
<td>Random decisioning</td>
</tr>
<tr>
<td>Reclassification with derogs</td>
<td>Simulate policy rules</td>
</tr>
<tr>
<td>Reclassification with score</td>
<td>Extreme rejects by score</td>
</tr>
<tr>
<td>Re-weighting</td>
<td>Accept / reject overlap</td>
</tr>
</tbody>
</table>
## When to use each method

<table>
<thead>
<tr>
<th>Method</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>Low reject rate</td>
</tr>
<tr>
<td></td>
<td>Random decisioning</td>
</tr>
<tr>
<td>Reclassification with derogs</td>
<td>Simulate policy rules</td>
</tr>
<tr>
<td>Reclassification with score</td>
<td>Extreme rejects by score</td>
</tr>
<tr>
<td>Re-weighting</td>
<td>Accept / reject overlap</td>
</tr>
<tr>
<td>Parcelling</td>
<td>Strict adherence to scorecard</td>
</tr>
</tbody>
</table>
## When to use each method

<table>
<thead>
<tr>
<th>Method</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>Low reject rate</td>
</tr>
<tr>
<td></td>
<td>Random decisioning</td>
</tr>
<tr>
<td>Reclassification with derogs</td>
<td>Simulate policy rules</td>
</tr>
<tr>
<td>Reclassification with score</td>
<td>Extreme rejects by score</td>
</tr>
<tr>
<td>Re-weighting</td>
<td>Accept / reject overlap</td>
</tr>
<tr>
<td>Parcelling</td>
<td>Strict adherence to scorecard</td>
</tr>
<tr>
<td>Heckman</td>
<td>To test bias</td>
</tr>
<tr>
<td></td>
<td>Easy to use</td>
</tr>
<tr>
<td>Method</td>
<td>Use</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>None</td>
<td>Low reject rate</td>
</tr>
<tr>
<td></td>
<td>Random decisioning</td>
</tr>
<tr>
<td>Reclassification with derogs</td>
<td>Simulate policy rules</td>
</tr>
<tr>
<td>Reclassification with score</td>
<td>Extreme rejects by score</td>
</tr>
<tr>
<td>Re-weighting</td>
<td>Accept / reject overlap</td>
</tr>
<tr>
<td>Parceling</td>
<td>Strict adherence to scorecard</td>
</tr>
<tr>
<td>Heckman</td>
<td>To test bias</td>
</tr>
<tr>
<td></td>
<td>Easy to use</td>
</tr>
<tr>
<td>Bureau</td>
<td>Quality bureau match</td>
</tr>
</tbody>
</table>
Combinations of approaches!

- Sometimes essential
- Depends on technique used
- Depends on past decisioning
- Depends on sample available
Conclusions

- Need for reject inference influenced by decline rate
- All methods discussed are valid under assumptions
- However, the best method varies case to case and a method may be invalid if assumptions are violated
- Select method according to the portfolio and validity of assumptions
- Use real outcome information on accounts if available
- Frequently require multiple approaches