MAPA Mapping

Scorecard Calibration using a Monotone Adjacent Pooling Algorithm

Presented at
Edinburgh Credit Scoring and Control IX
September 7-9, 2005

Raymond Anderson
Standard Bank Group
Johannesburg, South Africa

Standard Bank

Why Calibrate?

- **Traditional Retail Norm**
  - Focus on ranking ability
  - Initially an Accept/Reject paradigm
- **Consistent meaning**
  - Across Scorecards
  - Over Time
  - Across Products
- **Moving Forward**
  - Focus on predictive accuracy
  - Pricing
  - Provisioning
  - Capital Adequacy

On What Measure?

- **Good/Bad Definition**
  - Bad = 60 days past due
  - Good = Current
  - Indet = Mid Range
- **Basel II**
  - Bad=90 days past due
  - Good=Not Bad
  - No Mid Range

NOTE: Basel based upon corporate methodologies.

DEFAULT / NOT DEFAULT

Standard Bank

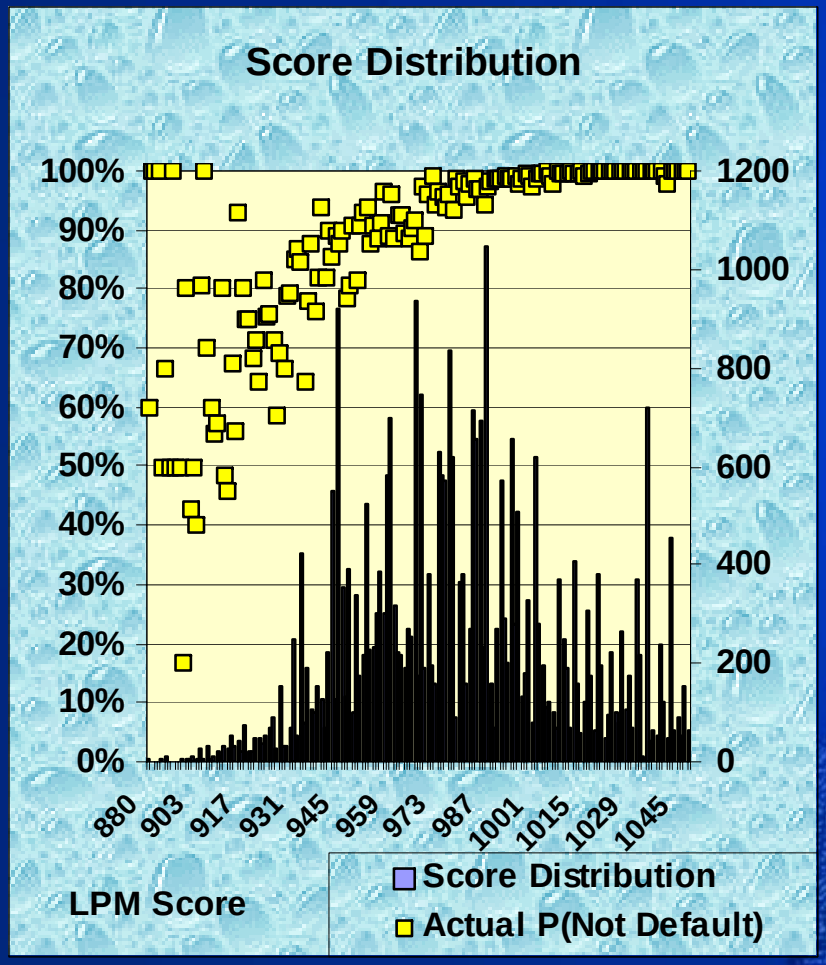
Predictive Modeling Techniques

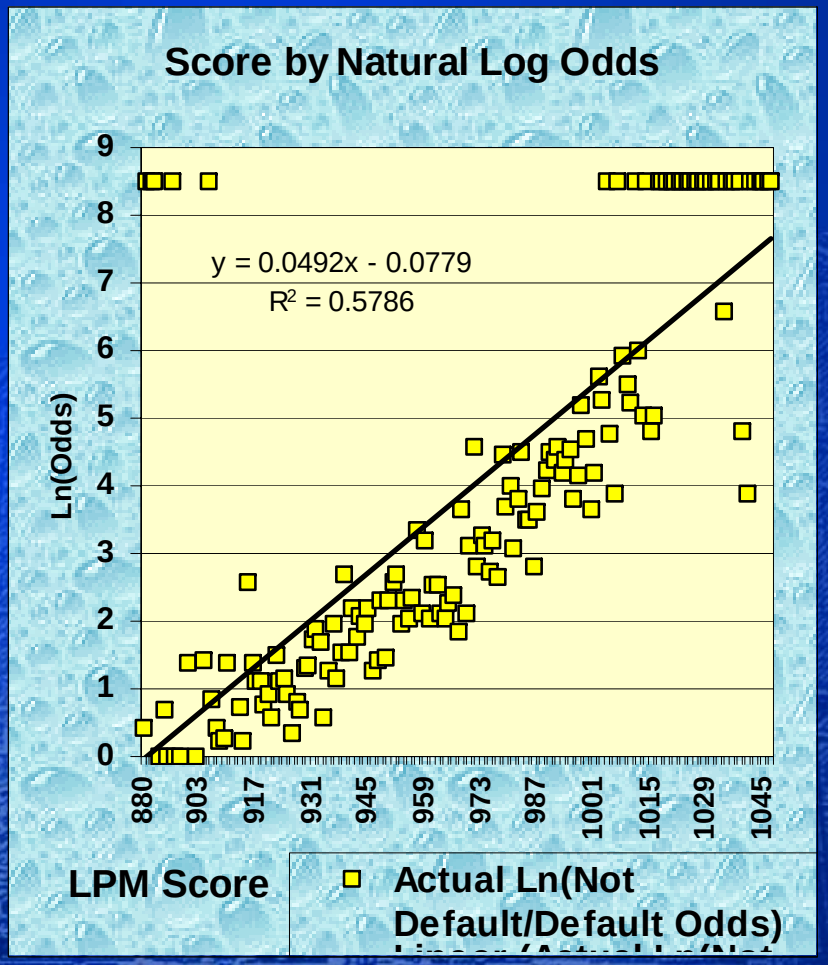
- Logistic Regression
  - Reasonable estimates, if Basel def'n used
- Linear Probability Modeling
  - Ranks well, but scores unreliable as estimates
  - Generalised Additive Non-parametric Regression
- Decision Trees
  - Use historical results directly

If Basel definition not used, or probability estimates unreliable, then mapping is necessary.

Standard Bank

LPM Score Results





Standard Bank

Gini = 55.52%

Assumptions

- **Ranking Ability paramount!**
- **Estimates necessary, but secondary**

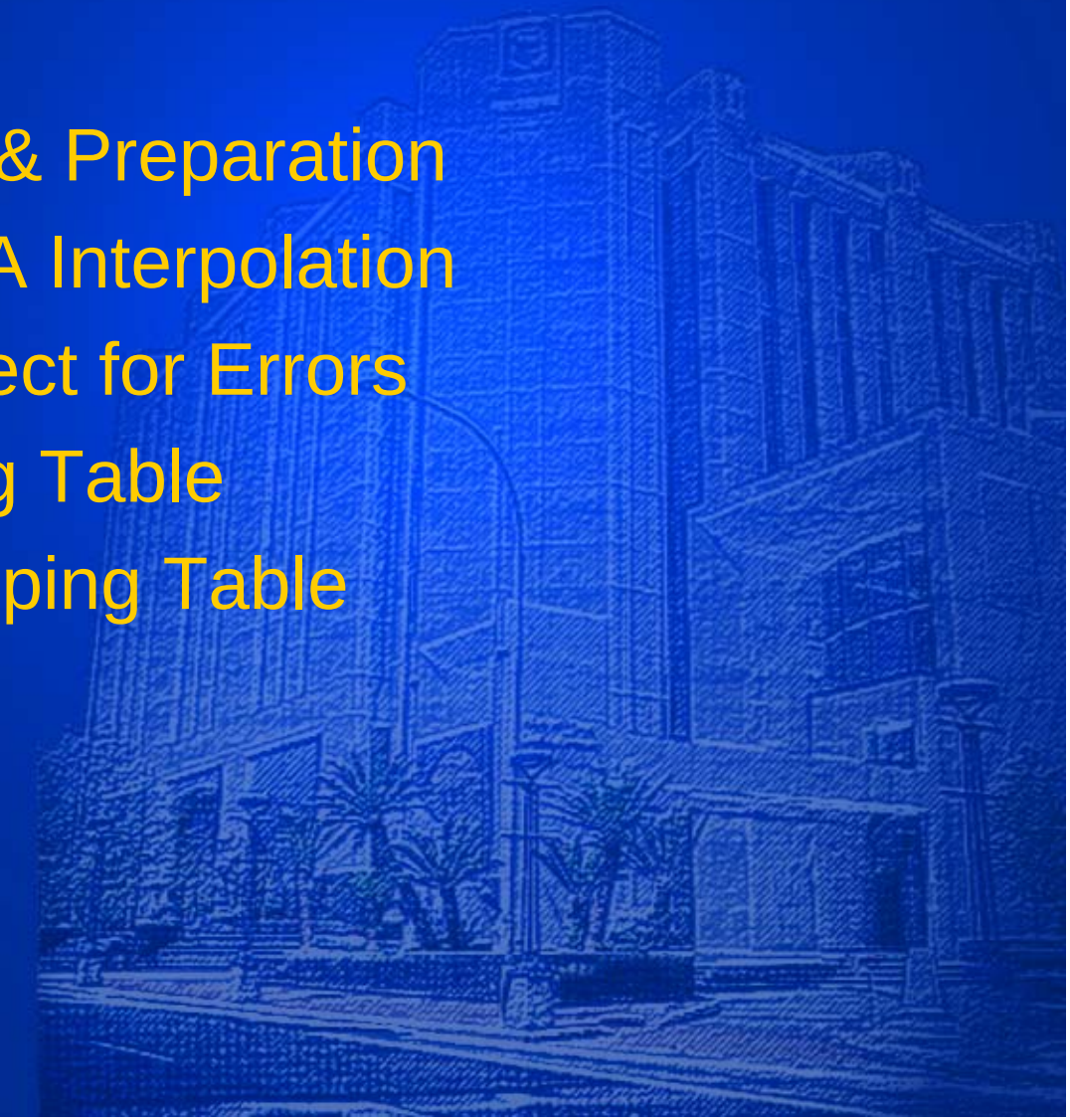
Possible Methodologies

- **Score banding / Risk Indicators**
  - **Use Historical Figures**
  - **Grouping Unscientific**
- **Logit, with score as sole independent**
  - **Simple, but assumes linearity**
- **Fitting of Lorenz curve (Glößner 2003)**
  - **Very complicated**

Standard Bank

MAPA Mapping - Process

I. Data Selection & Preparation
II. 1st Pass: MAPA Interpolation
III. 2nd Pass: Correct for Errors
IV. Create Mapping Table
V. Implement Mapping Table



Standard Bank

Data Selection & Preparation

A. **Out of time/out of sample?**
B. **Within Universe**
C. **Rank by Score**
D. **Set Target Variable**

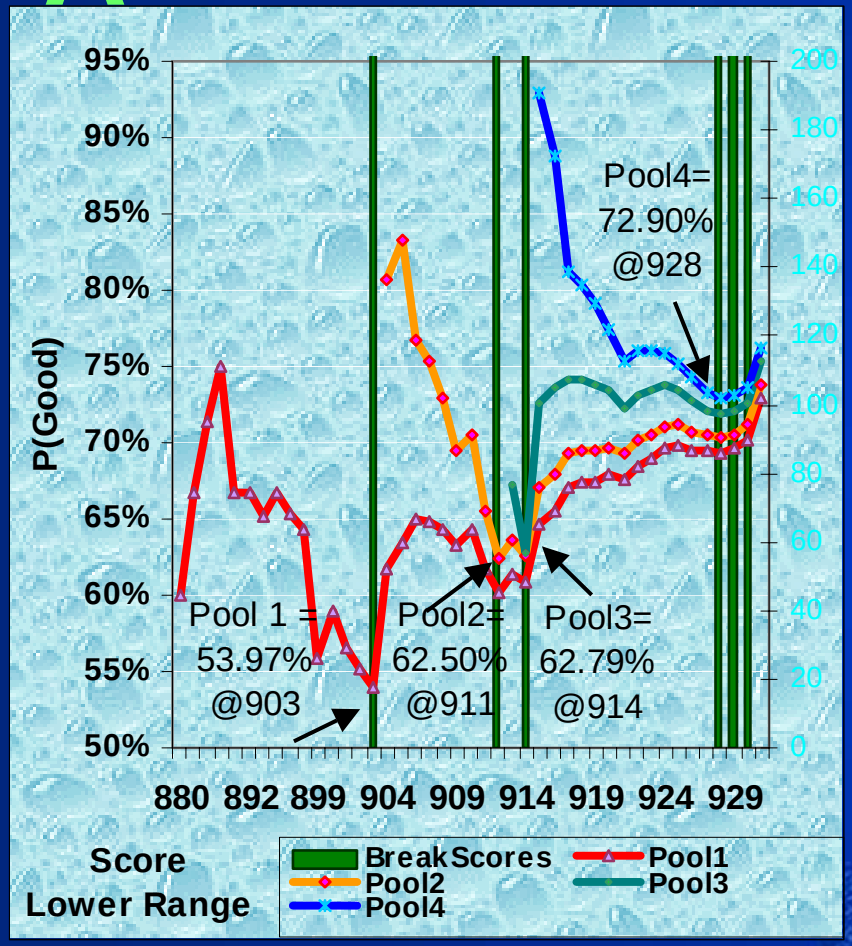
1st Pass: MAPA Interpolation

A. **Apply MAPA to identify Pools**
B. **Calculate Ln(odds) per Pool**
C. **Interpolate High and Low Ln(Odds) for each Pool**
D. **Interpolate Ln(Odds) for each Record**

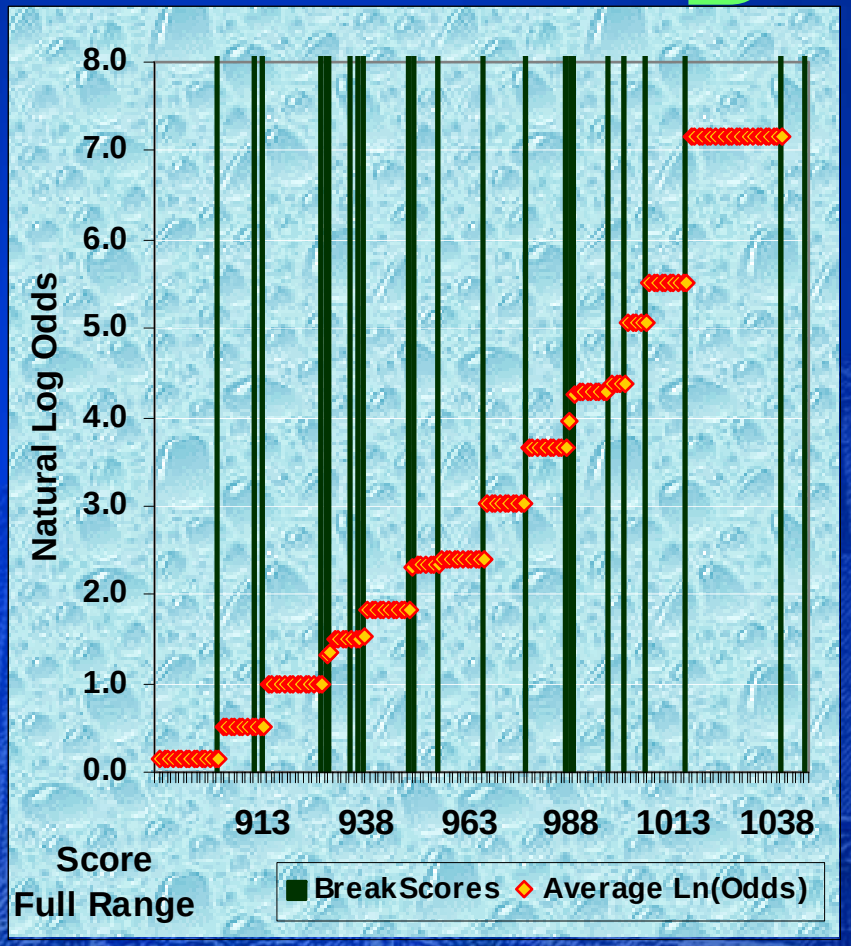
Standard Bank

Pool Definition

A



B

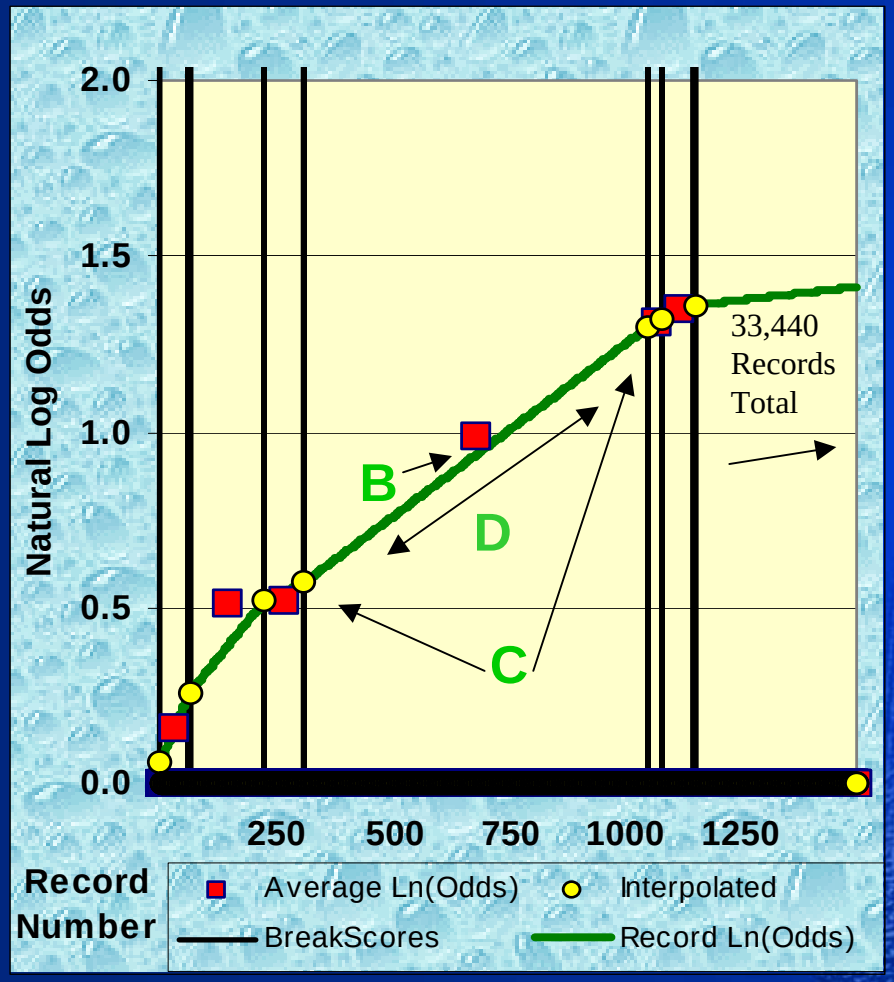


If monotonic, then P(Good) increases with score. Use Iterative process: a) find score with lowest cumulative P(Good); b) set that score as upper bound for pool; c) clear and repeat with remaining scores until all scores pooled.

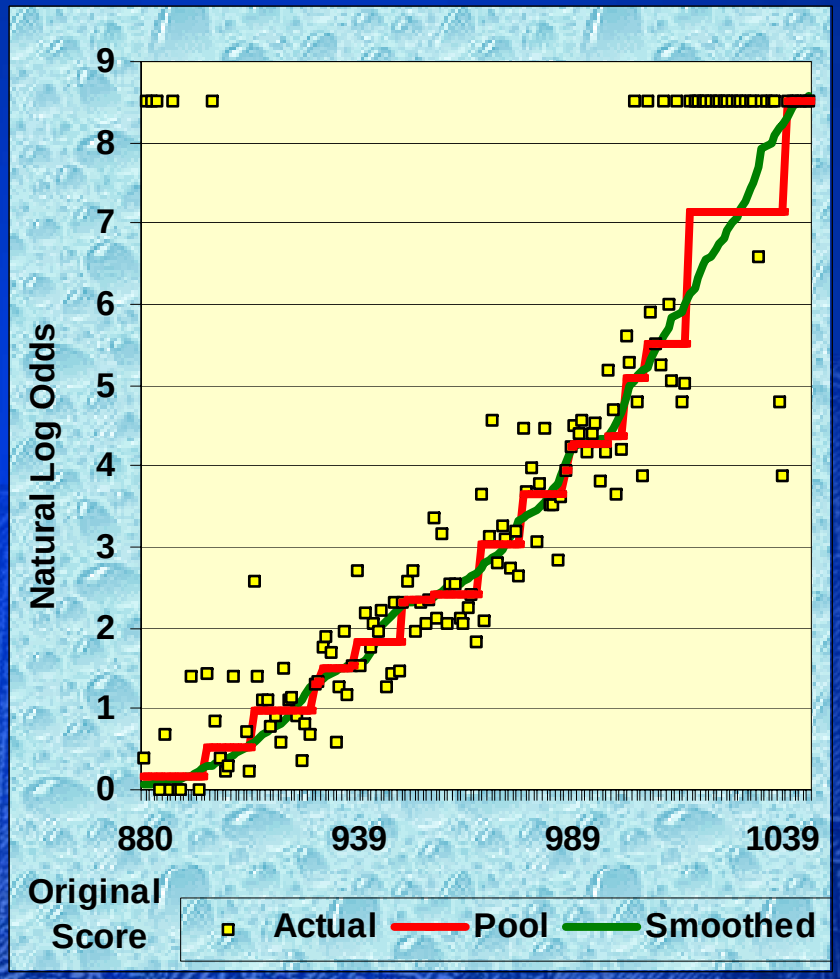
Gini = 56.22%

Standard Bank

Interpolation



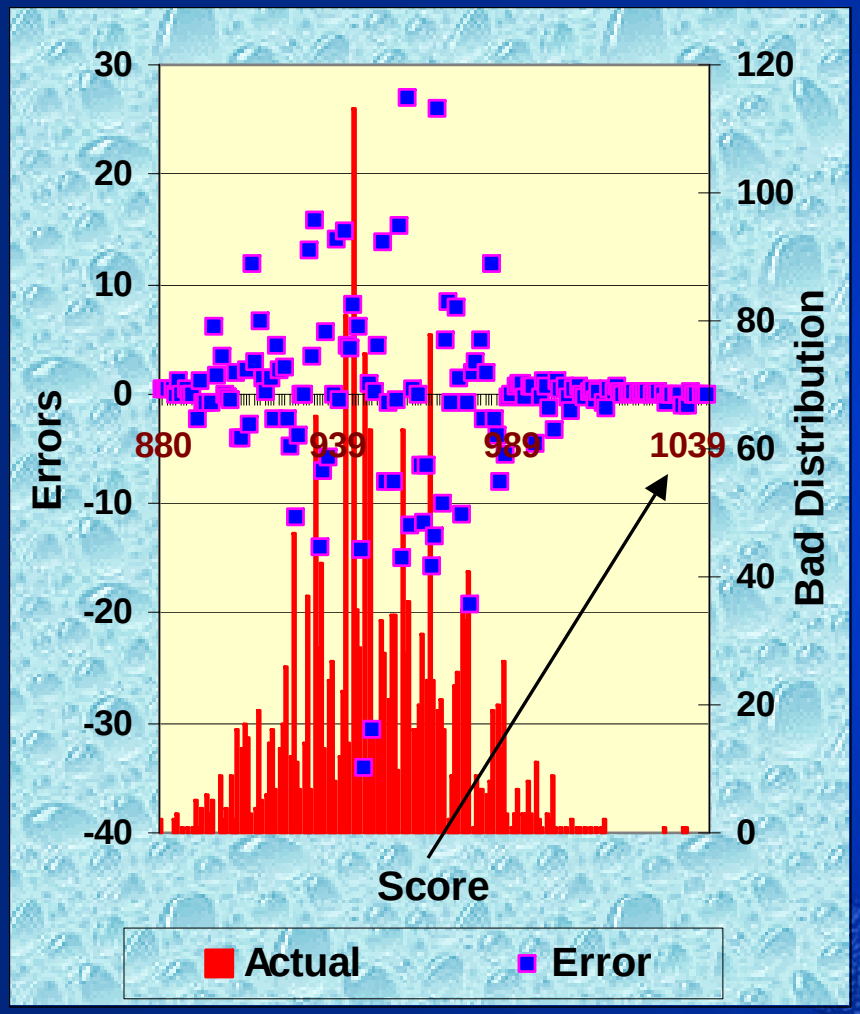
1st Pass Results

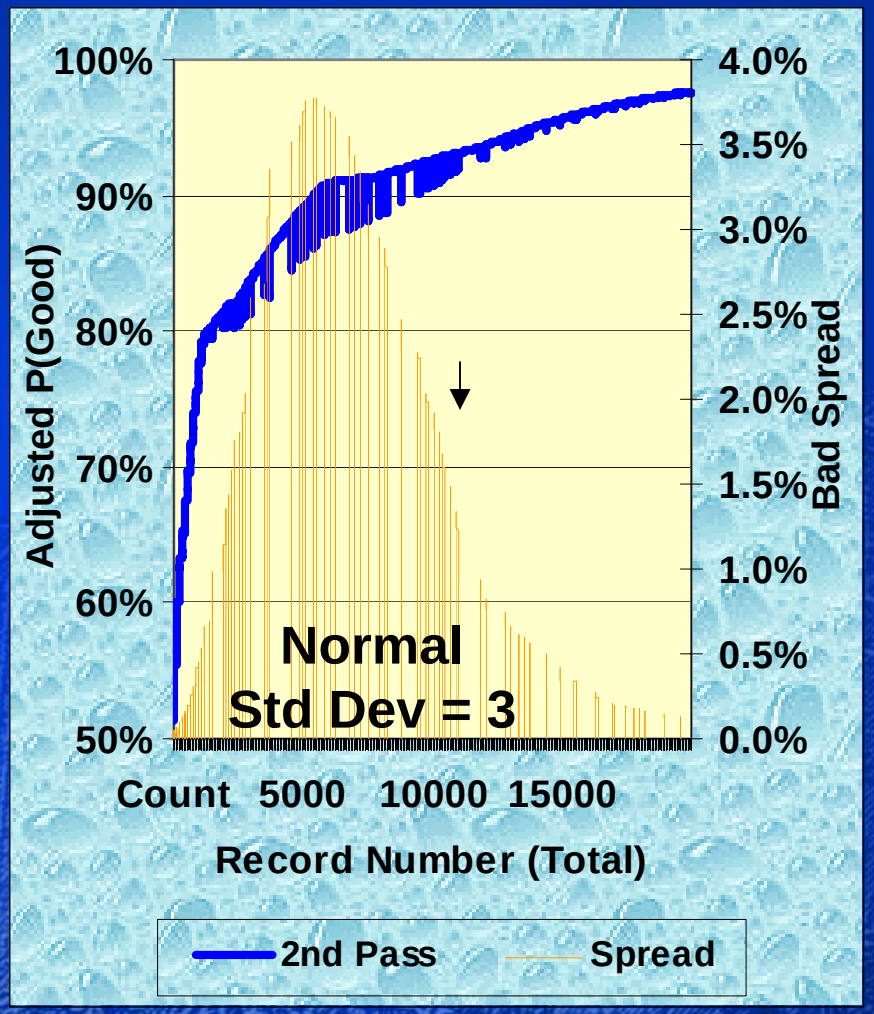


Use average P(Good) per Pool [B] to Interpolate Ln(Odds) for breakrecords [C], and use to interpolate Ln(Odds) for all records [D]. Aggregate by score, and we have a smoothed score to Ln(Odds) mapping… but with errors.

Standard Bank

2nd Pass: Error Correction

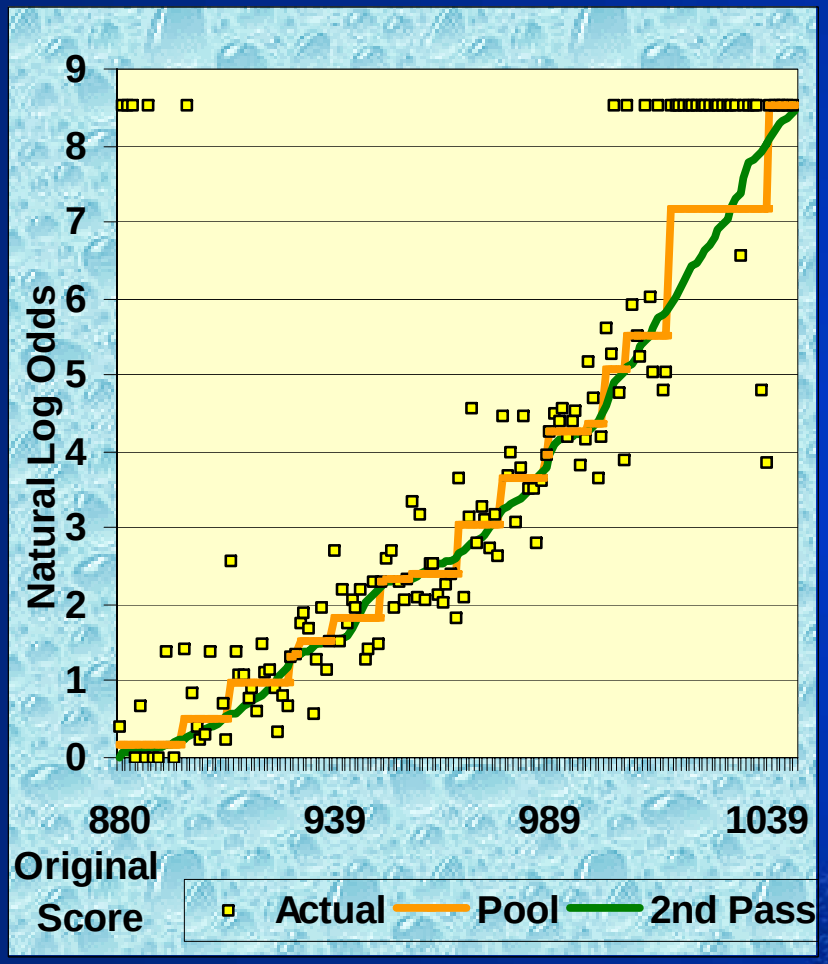


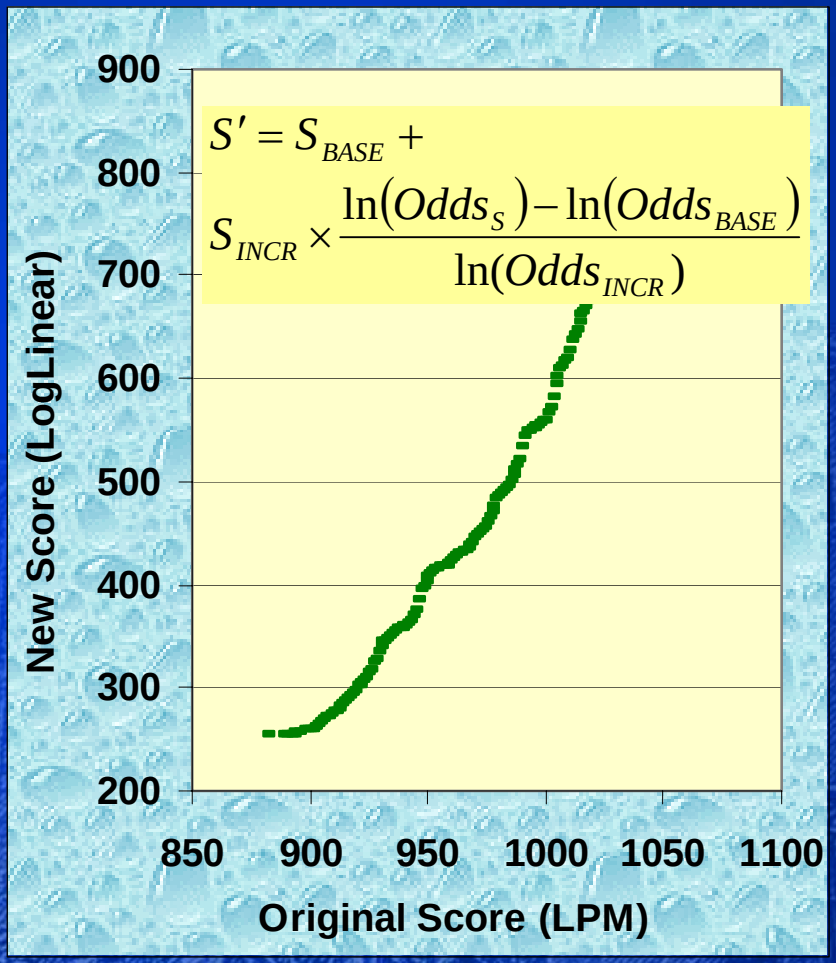


Net deficiency of 31.3 bads out of 1,990 (1.6%), distributed in same fashion as bads. Error corrected by spreading over bads, assuming normally distributed with Z-value from –3 to +3. P(Good) estimate adjusted downwards.

Standard Bank

Step 3: Mapping Table





Now have final P(Good). We can map Ln(Odds) onto new loglinear scores. Example at right has 32/1 odds at baseline 500, doubling every fifty points.

Gini = 55.51%

Standard Bank

Rescale

$$S' = S_{BASE} + \frac{S_{INCR}}{\ln(Odds_{INCR})} \times \left(\ln(Odds_S) - \ln(Odds_{BASE})\right)$$

Base Score = 500

Base Odds = 32

Double Every 50 points

$$(Odds \mid S = 989) = 60.47$$

$$S'_{989} = 500 + \frac{50}{\ln(2)} \times \left(\ln(60.47) - \ln(32)\right) = 545.9$$

Thus, score of 989 maps to 546.
Converts to Bad Rate of 1.625%

$$P(Bad) = \left(1 + \exp\left((S' - S_{BASE}) \times \frac{\ln(Odds_{INCR})}{S_{INCR}} + \ln(Odds_{BASE})\right)\right)^{-1}$$

$$P(Bad \mid S' = 546) = 1.625\% =$$

$$\left(1 + \exp\left((546 - 500) \times \frac{\ln(2)}{50} + \ln(32)\right)\right)^{-1}$$

Standard Bank

Conclusions

- Historical focus on ranking ability (power)
- Need reasonable estimates (accuracy)
- Problems where scorecard build and required definitions differ, where estimates unreliable, or significant changes to business environment..
- Requirement:
  - Business issues drive scorecard development
  - Apply transformations to obtain PD estimates for Basel II

Standard Bank

Conclusion cont'd

ADVANTAGES

- Conceptually Simple
- Non-Linear
- No Power Loss
- Handles any Binary Transformation
- Allows updates using latest performance
  - Historical (Detailed)
  - Informed (Constant)

ISSUES

- Always Backward Looking!!!
- Requires Mapping Table
- Small Numbers? Bias?
- Raw scores still needed
  - Scorecard Monitoring
  - Strategy???
- Endpoint Treatment?
- Other variations may provide improvements

Standard Bank