Actual = Expected:
Statistical Framework for Scorecard Management

Scoreplus Webinar
31 March 2020

Gerard Scallan
gerard.scallan@scoreplus.com
Actual = Expected:
Statistical Framework for Scorecard Management

- Sufficient Statistic
- Distance and Certainty
- Scorecard Development
- Scorecard Monitoring
- Coronavirus Crisis
Complexity: the Enemy

Complexity → Confusion → Unmanageable → Wrong Decisions
Goal: Simplify Policies

Dimensionality Reduction → Scores

DEFINITION:
Score = Sufficient Statistic for Aspect of Behaviour

Sufficient Statistic:
“No other statistic which can be calculated from the same sample provides any additional information as to the value of the parameter”

- R.A. Fisher (1922) p. 310

E.g. if Pr(Default) = 5%, then must have
Pr(Default | Own) = Pr(Default | Rent) = 5%

e.g. Parameter = PD
Sufficient Statistic ↔ Maximum Likelihood

- Find scorecard values which maximise:
  - $\Pr(\text{Good})$ for the goods in sample
  - $\Pr(\text{Bad})$ for the bads in sample

Maximise estimated probabilities of actual outcomes
Maximum Likelihood + LogOdds Model → Logistic Regression

- Maximum Likelihood
  - Log Odds
  - → Logistic Regression
  - → Actual = Expected
    - For scorecard build sample
    - Proof: See Appendix

Default by Loan Term

Operational Definition of Sufficient Statistic
What is “Expected”? 

Model implies “expected” outcome for each sample point
Characteristic in model

*(Categorical variables)*

<table>
<thead>
<tr>
<th>Attr. Group</th>
<th>Goods Count</th>
<th>Bads Count</th>
<th>Total Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;= 30 m</td>
<td>2690</td>
<td>489</td>
<td>3179</td>
</tr>
<tr>
<td>31-84 m</td>
<td>2575</td>
<td>223</td>
<td>2798</td>
</tr>
<tr>
<td>85+ m</td>
<td>1936</td>
<td>181</td>
<td>2117</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>7201</strong></td>
<td><strong>893</strong></td>
<td><strong>8094</strong></td>
</tr>
</tbody>
</table>

**EXPECTED (by score)**

<table>
<thead>
<tr>
<th>Goods Count</th>
<th>Bads Count</th>
<th>Total Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>2690.0</td>
<td>489.0</td>
<td>3179.0</td>
</tr>
<tr>
<td>2575.0</td>
<td>223.0</td>
<td>2798.0</td>
</tr>
<tr>
<td>1936.0</td>
<td>181.0</td>
<td>2117.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>7201.0</strong></td>
<td><strong>893.0</strong></td>
</tr>
</tbody>
</table>

**Exact Equality → Maximum Likelihood Equations**
Actual = Expected: Statistical Framework for Scorecard Management

- ✔ Sufficient Statistic
- → Distance and Certainty
- ❏ Scorecard Development
- ❏ Scorecard Monitoring
- ❏ Coronavirus Crisis
## Would Predictor Add Value to Model?

### Marginal Chi² Test Measures Certainty

- **Null Hypothesis**: Existing score accurately estimates probabilities
  - Probabilities generate “expected” values in each cell

<table>
<thead>
<tr>
<th>Debit Turnover</th>
<th>OBSERVED</th>
<th>EXPECTED</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Goods</td>
<td>Bads</td>
</tr>
<tr>
<td>&lt;= 1000</td>
<td>436</td>
<td>174</td>
</tr>
<tr>
<td>&lt;= 3500</td>
<td>525</td>
<td>101</td>
</tr>
<tr>
<td>&gt; 3500</td>
<td>6240</td>
<td>618</td>
</tr>
<tr>
<td>Total</td>
<td>7201</td>
<td>893</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Goods</th>
<th>Bads</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;= 1000</td>
<td>487.7</td>
<td>122.3</td>
<td>610</td>
</tr>
<tr>
<td>&lt;= 3500</td>
<td>533.9</td>
<td>92.1</td>
<td>626</td>
</tr>
<tr>
<td>&gt; 3500</td>
<td>6179.4</td>
<td>678.6</td>
<td>6858</td>
</tr>
<tr>
<td>Total</td>
<td>7201</td>
<td>893</td>
<td>8094</td>
</tr>
</tbody>
</table>

- Calculated on model build sample:
  - Intercept term in model guarantees actual = expected for total sample
  - Use Log-Likelihood Chi² - a matter of taste!

<table>
<thead>
<tr>
<th>Observed pattern not explained by model estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>→ score is not a sufficient statistic for risk</td>
</tr>
</tbody>
</table>
Chi² Measure - Pros and Cons

Pros

- Identify candidates to enter model
- For many potential predictors, expected → actual rapidly
  - As terms added to model
  - Shows common information
  - Gives understanding of collinearity structure
- Highlights “incremental” information

Cons

- Many very significant misfits
- Chi² measures certainty
  - not distance
  - 0.0000009% vs. 0.0000007% meaningless
- Degrees of freedom ambiguous
  - Classed characteristics
- Chi² statistic proportional to sample size
  - Hinders learning across samples
- Beware of false positives!

Right idea – wrong packaging
### How Big is the Gap?

**Marginal Information and Delta Scores**

#### Weight of Evidence (WoE)

- \( \text{WoE} = \log \left( \frac{\text{Attribute Odds}}{\text{Population Odds}} \right) \)

#### Delta Score

- \( \text{Delta Score} = \text{Observed WoE} - \text{Expected WoE} \)

#### Approximation

- Approximation to score coefficients needed to line up expected with observed scores.

#### Marginal Information Value

- \( \text{Marginal Information Value} = \text{Avg}_{\text{Good}}(\text{Delta Score}) - \text{Avg}_{\text{Bad}}(\text{Delta Score}) \)

- Similar to Kullback-Liebler Information Value

- Increased spread between average score of goods and bads

- ... if this characteristic brought into model

#### Table

<table>
<thead>
<tr>
<th>Debit Turnover</th>
<th>OBSERVED</th>
<th>EXPECTED</th>
<th>( \Delta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goods</td>
<td>Bads</td>
<td>Total</td>
<td>Goods</td>
</tr>
<tr>
<td>&lt;= 1000</td>
<td>436</td>
<td>174</td>
<td>-1.169</td>
</tr>
<tr>
<td>&lt;= 3500</td>
<td>525</td>
<td>101</td>
<td>-0.439</td>
</tr>
<tr>
<td>&gt; 3500</td>
<td>6240</td>
<td>618</td>
<td>0.225</td>
</tr>
<tr>
<td>Total</td>
<td>7201</td>
<td>893</td>
<td>0.000</td>
</tr>
</tbody>
</table>

- \( \chi^2 = 32.14 \)

- D.F. = 2

- p-value = 0.00001%

**Marginal Information Value** = 0.085
Measuring Collinearity

Overlaps in predictive power

- Most information is not unique to a single characteristic
- Delta scores reduce in magnitude as “correlated” variables enter model

Small Delta Scores → Information already covered by other characteristics in model
**End Game: Actual ≈ Expected Everywhere**

<table>
<thead>
<tr>
<th>Debit Turnover</th>
<th>OBSERVED</th>
<th>EXPECTED</th>
<th>Δ</th>
<th>SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goods Bads Total</td>
<td>Goods</td>
<td>Bads</td>
<td>Total</td>
<td>Goods</td>
</tr>
<tr>
<td>&lt;= 1000</td>
<td>436</td>
<td>174</td>
<td>-1.169</td>
<td>447.3</td>
</tr>
<tr>
<td>&lt;= 3500</td>
<td>525</td>
<td>101</td>
<td>-0.439</td>
<td>527.3</td>
</tr>
<tr>
<td>&gt; 3500</td>
<td>6240</td>
<td>618</td>
<td>0.225</td>
<td>6226</td>
</tr>
<tr>
<td>Total</td>
<td>7201</td>
<td>893</td>
<td>0.000</td>
<td>7201</td>
</tr>
</tbody>
</table>

\[ \text{Chi}^2 = 1.44 \quad \text{D.F.} = 2 \quad \text{p-value} = 48.60123\% \]

Marginal Information Value = 0.018

- Actual (approximately) = Expected
- Otherwise characteristic would come into model

**End Point: Actual ≈ Expected on all characteristics... whether in model or not!**
Measurement Framework

*Gap Between Reality and Model*

- **Distance:** Marginal Information Value
  - How big is the improvement expected from bringing this predictor into model?

- **Score Corrections:** Delta Scores
  - What would score coefficients look like?

- **Certainty:** Marginal Chi²
  - Is sample big enough for results to be reliable?
Actual = Expected: Statistical Framework for Scorecard Management

- Sufficient Statistic
- Distance and Certainty
- Scorecard Development
- Scorecard Monitoring
- Coronavirus Crisis
Applications of Marginal Analysis

Scorecard Development

Model Development

1. Select characteristics to enter model
2. Model segmentation
3. Policy rule assessment

Model Validation

4. Identify reasons for scorecard deterioration
   ... and correct them
5. Business source evaluation
6. Comprehensive Credit Reporting

Swiss Army Knife

Used “Everywhere” in Scorecard Management
### Marginal Analysis Application 1

#### Selecting Scorecard Characteristics

At each step...
- Rank predictors by (Marginal) IV
- Predictor with maximum Marginal IV enters model
- ... provided Marginal Chi² can be “made” significant
- Continue until no **significant** Marginal IV left
- Significance Threshold: MIV > .020
  - Marginal Chi² p-level < 5%

**Zero Marginal Information = Sufficient Statistic**

#### Tabela de Valores

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>IV</th>
<th>Score1</th>
<th>Score2</th>
<th>Score3</th>
<th>Score4</th>
<th>Score5</th>
</tr>
</thead>
<tbody>
<tr>
<td>CurBal</td>
<td>0.032</td>
<td>0.019</td>
<td>0.017</td>
<td>0.013</td>
<td>0.010</td>
<td>0.008</td>
</tr>
<tr>
<td>CurCTO</td>
<td>0.185</td>
<td>0.121</td>
<td>0.086</td>
<td>0.089</td>
<td>0.007</td>
<td>0.006</td>
</tr>
<tr>
<td>CurDaysXs</td>
<td>0.616</td>
<td>0.125</td>
<td>0.113</td>
<td><strong>0.106</strong></td>
<td><strong>0.094</strong></td>
<td>0.021</td>
</tr>
<tr>
<td>CurDTO</td>
<td>0.215</td>
<td>0.117</td>
<td>0.087</td>
<td>0.093</td>
<td>0.026</td>
<td>0.025</td>
</tr>
<tr>
<td>CurValXs</td>
<td>0.515</td>
<td>0.121</td>
<td>0.110</td>
<td>0.093</td>
<td>0.090</td>
<td>0.007</td>
</tr>
<tr>
<td>ToB</td>
<td>0.692</td>
<td>0.526</td>
<td>0.010</td>
<td>0.026</td>
<td>0.025</td>
<td>0.025</td>
</tr>
<tr>
<td>MthsInact</td>
<td>0.012</td>
<td>0.005</td>
<td>0.001</td>
<td>0.004</td>
<td>-0.002</td>
<td>-0.003</td>
</tr>
<tr>
<td>MthsNoCTO</td>
<td>0.077</td>
<td>0.066</td>
<td>0.043</td>
<td>0.045</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>NetTO</td>
<td>0.074</td>
<td>0.028</td>
<td>0.007</td>
<td>0.010</td>
<td>0.002</td>
<td>0.000</td>
</tr>
<tr>
<td>DaysDbL3m</td>
<td>0.055</td>
<td>0.008</td>
<td>0.013</td>
<td>0.008</td>
<td>0.005</td>
<td>0.004</td>
</tr>
<tr>
<td>DaysXsL6m</td>
<td><strong>0.856</strong></td>
<td>0.000</td>
<td>0.008</td>
<td>0.011</td>
<td>0.015</td>
<td>0.012</td>
</tr>
<tr>
<td>CurMxBal</td>
<td>0.033</td>
<td>0.015</td>
<td>0.018</td>
<td>0.013</td>
<td>0.005</td>
<td>0.003</td>
</tr>
<tr>
<td>DishL1m</td>
<td>0.291</td>
<td>0.090</td>
<td>0.084</td>
<td>-0.006</td>
<td>-0.008</td>
<td>-0.010</td>
</tr>
<tr>
<td>DishL3m</td>
<td>0.292</td>
<td>0.081</td>
<td>0.077</td>
<td>0.005</td>
<td>0.011</td>
<td>0.011</td>
</tr>
<tr>
<td>SinceDish</td>
<td>0.810</td>
<td>0.397</td>
<td><strong>0.299</strong></td>
<td>0.057</td>
<td>0.050</td>
<td><strong>0.051</strong></td>
</tr>
<tr>
<td>InterCTO</td>
<td>0.017</td>
<td>0.004</td>
<td>-0.003</td>
<td>-0.004</td>
<td>-0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td>InterDTO</td>
<td>0.003</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>-0.002</td>
<td>-0.002</td>
</tr>
<tr>
<td>AutoCr</td>
<td>0.209</td>
<td>0.143</td>
<td>0.108</td>
<td><strong>0.106</strong></td>
<td>0.005</td>
<td>0.004</td>
</tr>
<tr>
<td>ValDishL6m</td>
<td>0.468</td>
<td>0.145</td>
<td>0.137</td>
<td>-0.001</td>
<td>-0.001</td>
<td>0.003</td>
</tr>
</tbody>
</table>
Application 2: Model Segmentation

Interactions

- Current practice: split if different information available
  - E.g. Limited companies vs. sole traders in SMEs
  - “Cleans” vs. “Dirties” on credit card behavioural score
  - Specific characteristics only to apply to sub-population

- But each characteristic has a “not applicable” category
  - E.g. Sole Traders
  - With Actual = Expected on Cleans

- So presence of “cleans” has no influence on “dirty” characteristics

- Real question:
  - **DO COMMON PREDICTORS HAVE DIFFERENT PATTERNS ON SEGMENTS?**

Interactions: Actual ≠ Expected
### Additional Information in Single Scorecard

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Shareholders' Funds</th>
<th>Delta IV</th>
<th>DF</th>
<th>Total</th>
<th>Ltd Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Ltd Co.</td>
<td>+0.116 -0.214</td>
<td>3</td>
<td>2</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Negative</td>
<td>-0.319 -0.145</td>
<td>-0.116</td>
<td>0.598</td>
<td>6.10%</td>
<td>2.51%</td>
</tr>
<tr>
<td>&lt; 100k</td>
<td>-0.411 -0.347</td>
<td>-0.064</td>
<td>7.37</td>
<td>7.37</td>
<td>7.37</td>
</tr>
<tr>
<td>100k +</td>
<td>0.936 0.338</td>
<td>0.598</td>
<td>4.64</td>
<td>4.64</td>
<td>4.64</td>
</tr>
<tr>
<td>TOTAL</td>
<td>0.000 0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

- Scorecard partially built
- Missing: Actual ≠ Expected
  - Ltd. Company already in model
  - Actual ≠ Expected on other atts.
  - But differences not significant?

**Extra Information**

→ Extra Characteristics
≠ Not Extra Scorecard
→ Fewer Splits Needed
Model Segmentation

Testing for Interactions

- Characteristic interactions
  - Multiple models
    - e.g. Delinquency - Time on books
- Test for Actual = Expected on each subpopulation
  - For each predictive characteristic
  - Systematic screen for interactions
- Small samples => Statistics matter!
- Shows many splits unnecessary

Clear conceptual framework (and algorithm) for tough problem
### Application 3: Policy Rule Analysis

**Turn Down if Bad Credit Bureau?**

- Recovery and Bad Debt groups previously subject to policy rule
  - In principle all rejected – but some exceptions
  - Exceptions are exceptional!

- Analyse interaction with score
- Also look at future accept rates
- ... and values

#### Performance well explained by model

<table>
<thead>
<tr>
<th>Status</th>
<th>OBSERVED</th>
<th>EXPECTED</th>
<th>Δ SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Goods</td>
<td>Bads</td>
<td>Total</td>
</tr>
<tr>
<td>Recovery</td>
<td>91</td>
<td>59</td>
<td>-3.256</td>
</tr>
<tr>
<td>Bad Debt</td>
<td>7</td>
<td>45</td>
<td>-5.550</td>
</tr>
<tr>
<td>All Other</td>
<td>99902</td>
<td>2396</td>
<td>0.042</td>
</tr>
<tr>
<td>Total</td>
<td>100000</td>
<td>2500</td>
<td>0.000</td>
</tr>
</tbody>
</table>

- Chi² = 7.54  
- D.F. = 2  
- p-value = 2.30323%  
- Marginal Information Value = 0.016

- Chi² significant = Reliable
- But –ve Δ-scores (-9.6 on 20 PDO scale)

---

**Goal: Simplify Policy ➔ Reduce Complexity**

---
Actual = Expected:
Statistical Framework for Scorecard Management

✓ Sufficient Statistic
✓ Distance and Certainty
✓ Scorecard Development
→ Scorecard Monitoring
✓ Coronavirus Crisis
Tracking Approach (and model validation!)

Score = Log (Odds)

Validation Process
- Key business decisions assume score-risk relationship
  - Basis for strategies
  - Needs management assumptions on PiT parameters
- Fit logistic regression on validation population
  - Evaluates overall performance of model
  - Ensures Actual = Expected for total population
- Starting point for Marginal Analysis

Marginal Analysis reports should be part of regular monitoring
Application 4: Scorecard Deterioration

Why is the model not working properly?

Actual ≠ Expected
→ Behaviour patterns changed
→ Model is wrong OR
→ Strategy is wrong OR BOTH

Assumptions underlying policies are wrong

Corrective Action
□ One characteristic is off:
□ Apply Δ-scores – quick correction
□ Several characteristics are off:
□ Mother-child scorecard
□ Generalisation of Δ-scores

Goal: Dynamic Scores – change as behaviour patterns change
Change in Behaviour Patterns

Example of tracking analysis

- Clear WoE pattern
- IV: 0.616  Marginal IV: 0.072
  - But some negative contributions
- The Δ-scores show that scorecard “exaggerates”
  - Worst not as bad as scores suggest
- Why? Change in treatment of Excess?
- Zero excess (2/3 of population) is under-rated

Must explain the business story – statistics are not enough
Application 5: Business Source Evaluation

Evaluate Potential and Performance

ACCOUNT OPENINGS 2012/Q2

<table>
<thead>
<tr>
<th>Recruitment Channel</th>
<th>Default Rate</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Budget</td>
<td>At Opening</td>
</tr>
<tr>
<td>Direct Mail</td>
<td>3.5%</td>
<td>3.7%</td>
</tr>
<tr>
<td>Internet</td>
<td>5.0%</td>
<td>5.2%</td>
</tr>
<tr>
<td>Partners</td>
<td>4.0%</td>
<td>3.9%</td>
</tr>
<tr>
<td>Branch</td>
<td>4.8%</td>
<td>5.0%</td>
</tr>
<tr>
<td>Total</td>
<td>4.1%</td>
<td>4.2%</td>
</tr>
</tbody>
</table>

- “At opening” figures derived from scores on account opening time
  - Profile of applicants different from budget expectations
- Isolate departures from expectations
  - Take account of differing potential
- Can be extended to policy rules, marketing campaigns, collections strategies, ...
Actual = Expected:
Statistical Framework for Scorecard Management

✓ ❑ Sufficient Statistic
✓ ❑ Distance and Certainty
✓ ❑ Scorecard Development
✓ ❑ Scorecard Monitoring
→ ❑ Coronavirus Crisis
Application 6: Coronavirus Crisis

Effect on Credit

**Economy**

- Jobs Losses
  - E.g. US - 10m in two weeks
  - → Reduced income
  - Especially marginal workers
- Uncertainty
  - Reduced demand
  - Especially for cars, houses
- $S^3$: short (?) sharp shock
  - Quick rebound?
- Run rate of economy -30%
- Deflation
  - Oil down 40% in month

**Credit**

- Lower application volume
- Less business investment
- Missed payments
  - And fewer cures in delinquency
- Higher provision
  - IFRS9 is point in time
- Higher losses
  - Recognised from 2020 Q3
- Stable capital ($\pm$)
  - Not fully point in time
  - Capital is front-loaded

**Longer term: Effect on confidence**
Phase I: No Data
(Educated Guesswork)

Application Score = Log (Odds)

Adjustment

- Expect worse risk at each score
- Offset score-odds line
- How much? Look at 2008-09
  - But 3 times faster ...
  - MANAGEMENT JUDGMENT
- Slope doesn’t change?
  - Relative risks unchanged
- Reinforce policy rules
  - stability - is income sustainable?
  - Ability to repay
  - Cash flow for SMEs

Modified score-odds ➔ Cutoff and Pricing ➔ Marketing and Budget
Phase II: Headline Delinquency
(after ~ 3 months)

- Phase I Assumption
- → assumed monthly defaults
  - Use 2 payments in arrears
  - Cf. Portfolio Management Analytics course
- Compare to Actuals
- When? 3-4 mos. unsecured
  - 6 mos. on secured
- Statistical test: can I believe it?
  - Small samples
- Update business assumptions

Expected Early Default Rates

Survival depends on speed of reaction

Adjustment

Scoreplus Webinar – 6 April 2020
Session CONF07 © ScorePlus SARL 2020
Phase III: Refine Models and Policy
(after ~ 100 recession bads)

Actual ≠ Expected
→ Behaviour patterns changed
→ Model is wrong OR
→ Strategy is wrong OR BOTH

Assumptions underlying policies are wrong

Corrective Action
☐ One characteristic is off:
☐ Apply $\Delta$-scores – quick correction

☐ Several characteristics are off:
☐ Mother-child scorecard
☐ Generalisation of $\Delta$-scores

Validation Process

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Global IV</th>
<th>MIV</th>
<th>$\chi^2$</th>
<th>DF</th>
<th>p-level</th>
</tr>
</thead>
<tbody>
<tr>
<td>CurBal</td>
<td>0.032</td>
<td>-0.017</td>
<td>4.291</td>
<td>1</td>
<td>3.83%</td>
</tr>
<tr>
<td>CurCTO</td>
<td>0.185</td>
<td>0.039</td>
<td>7.395</td>
<td>6</td>
<td>28.59%</td>
</tr>
<tr>
<td>CurDaysXs</td>
<td>0.616</td>
<td>0.072</td>
<td>13.591</td>
<td>6</td>
<td>3.46%</td>
</tr>
<tr>
<td>CurDTO</td>
<td>0.215</td>
<td>0.025</td>
<td>6.458</td>
<td>4</td>
<td>16.75%</td>
</tr>
<tr>
<td>CurValXs</td>
<td>0.515</td>
<td>0.045</td>
<td>12.153</td>
<td>7</td>
<td>9.56%</td>
</tr>
<tr>
<td>ToB</td>
<td>0.692</td>
<td>-0.019</td>
<td>6.219</td>
<td>7</td>
<td>51.44%</td>
</tr>
<tr>
<td>MthsInact</td>
<td>0.012</td>
<td>-0.003</td>
<td>0.675</td>
<td>1</td>
<td>41.13%</td>
</tr>
<tr>
<td>MthsNoCTO</td>
<td>0.077</td>
<td>0.006</td>
<td>1.217</td>
<td>2</td>
<td>54.42%</td>
</tr>
<tr>
<td>NetTO</td>
<td>0.074</td>
<td>0.009</td>
<td>0.319</td>
<td>3</td>
<td>95.64%</td>
</tr>
<tr>
<td>DaysDbL3m</td>
<td>0.055</td>
<td>0.016</td>
<td>2.159</td>
<td>2</td>
<td>33.98%</td>
</tr>
<tr>
<td>DaysXsL6m</td>
<td>0.856</td>
<td>0.027</td>
<td>11.278</td>
<td>9</td>
<td>25.71%</td>
</tr>
<tr>
<td>CurMxBal</td>
<td>0.033</td>
<td>0.003</td>
<td>0.268</td>
<td>2</td>
<td>87.46%</td>
</tr>
<tr>
<td>DishL1m</td>
<td>0.291</td>
<td>-0.010</td>
<td>1.398</td>
<td>2</td>
<td>49.71%</td>
</tr>
<tr>
<td>DishL3m</td>
<td>0.292</td>
<td>0.011</td>
<td>2.29</td>
<td>5</td>
<td>80.77%</td>
</tr>
<tr>
<td>SinceDish</td>
<td>0.810</td>
<td>0.025</td>
<td>13.703</td>
<td>8</td>
<td>8.98%</td>
</tr>
<tr>
<td>InterCTO</td>
<td>0.017</td>
<td>-0.001</td>
<td>0.421</td>
<td>2</td>
<td>81.02%</td>
</tr>
<tr>
<td>InterDTO</td>
<td>0.003</td>
<td>-0.002</td>
<td>1.1</td>
<td>1</td>
<td>29.43%</td>
</tr>
<tr>
<td>AutoCr</td>
<td>0.209</td>
<td>-0.013</td>
<td>7.37</td>
<td>4</td>
<td>11.76%</td>
</tr>
<tr>
<td>ValDishL6m</td>
<td>0.468</td>
<td>0.017</td>
<td>4.051</td>
<td>6</td>
<td>66.98%</td>
</tr>
</tbody>
</table>

Marginal Analysis - 12 months outcome

Keep models in line with best data-guided business judgment
## Marginal Analysis as Management Tool

### Model Development

- Easily analyse 1000s of predictors ... and interactions
- ... automate discretisation
- ... with reasonable computation

### Ongoing Management

- Identify changes in performance ... rapidly
- ... and suggest corrections to model
- Measure effect of policy variation e.g. price levels, marketing mix

### Operational Implementation of Sufficient Statistic ...

→ Simplify Management Control Process

- Coherent framework for model selection
  - Distance → Marginal Information
  - Certainty → Marginal Chi²

- Evaluate policy rules
- Measure consequences of overrides
The Management Link: Vigour and Rigour

Policy Management

Line Management

Statistical techniques

Scorecards Strategies

Operations

Tracking

Makes scoring models more transparent to ordinary people
APPENDIX: TECHNICAL STUFF

- Actual = Expected for discrete and continuous predictors
- Open Questions for Researchers
- References
Actual = Expected Equations
... equivalent to Maximum Likelihood

Problem: estimate scorecard $\beta$ from sample of Goods $(G)$ and Bads $(B)$

For case $i$: $\Pr_\beta (i \in G) = \frac{e^{x_i' \beta}}{1 + e^{x_i' \beta}}$  \hspace{1cm} $\Pr_\beta (i \in B) = \frac{1}{1 + e^{x_i' \beta}}$

Likelihood Function: $L(\beta) = \prod_{i \in G} \frac{e^{x_i' \beta}}{1 + e^{x_i' \beta}} \times \prod_{i \in B} \frac{1}{1 + e^{x_i' \beta}}$

$\ln L(\beta) = \sum_{i \in G} x_i' \beta - \sum_{i \in G \cup B} \ln(1 + e^{x_i' \beta})$

Maximise by setting partial derivatives w.r.t. each component $j$ of $\beta$ to zero:

$$\frac{\partial \ln L(\beta)}{\partial \beta_j} = \sum_{i \in G} x_{ij} - \sum_{i \in G \cup B} \frac{e^{x_i' \beta} x_{ij}}{1 + e^{x_i' \beta}} = \sum_{i \in G} x_{ij} - \sum_{i \in G \cup B} x_{ij} \Pr_\beta (i \in G) = 0$$

Let $x_{ij} = 1$ if $i$ is in category $A_j$, $x_{ij} = 0$ otherwise:

$$\|A_j \cap G\| = \sum_{i \in A} \Pr_\beta (i \in G)$$

Actual Goods = Expected Goods
Actual = Expected
Continuous Variables

\[ \sum_{i \in G} x_{ij} = \sum_{i \in G} \frac{e^{x_i \beta}}{1 + e^{x_i \beta}} \text{ for a continuous } x_{ij} \]

or \[ \sum_{i \in G} x_{ij} = \sum_{i \in G} \Pr(\beta(i \in G))x_{ij} \]

Divide both sides by total number of "goods", \( \|G\| \)

Average over Actuals = Average over Expecteds

\[ \text{Avg}_G(x_{ij}) = \frac{1}{\|G\|} \sum_{i \in G} \Pr(\beta(i \in G))x_{ij} \]

Because of the intercept term

\[ \|G\| = \sum_{i \in G} \Pr(\beta(i \in G)) \]

so \[ \text{Avg}_G(x_{ij}) = \frac{\sum_{i \in G} \Pr(\beta(i \in G))x_{ij}}{\sum_{i \in G} \Pr(\beta(i \in G))} \]

or \[ \text{Avg}_G(x_{ij}) = \text{Avg}_{EG}(x_{ij}) \]

where EG is the set of "expected goods"

Average over Actuals = Average over Expecteds
Questions?????

Researchers, Students ...

- Continuous predictors
  - Use Marginal Kolmogorov-Smirnov
- Probabilities not homogeneous
  - Is \( \chi^2 \) still robust?
  - cf. Spiegelhalter test
- \( \Delta \)-scores alternatives
  - 1st iteration of Newton-Raphson
- Variance of \( \Delta \)-scores
  - Variance of expected WoE?
  - Use of re-sampling techniques
- Translate from log-odds language to PDese
- Sequential testing
  - Information from consistency of results over time?
- Extend to models other than Logistic Regression
  - Balance and revenue models
  - Survival analysis

Some trivial – others not
References