EL validation DB unsecured Basel model

Jan Scholma / Maarten Terpstra
ING DB NL / CRO / CRM - Private Persons
CRC 2013
Characteristics of unsecured lending portfolio

- **Scope - customers**
  - Over 4.5 mln. customers

- **Scope – products**
  - Consumer loans
  - Credit cards
  - Overdraft limits
  - Current accounts with unauthorized balance

- **Scope – outstanding**
  - Large number of relatively small outstanding
The problem

- PD/EAD/LGD are developed separately
- PD at customer level
- EAD/LGD at facility level
  - Facility = current account + related credit products
- Combined in EL calculation
- EL = PD * EAD * LGD
- Independence assumption
- True?
- Effect on EL?

Validate!
Simulation example

“A good simulation gives more insight than a bad research”
Theoretical example: the portfolio

- 300 customers
- 1 facility per customer
- OS per facility constant = 50
- Limit per facility constant = 100
- 2 PD classes, 2 EAD classes, 2 LGD classes
- Total limit = 30,000
- Total OS = 15,000
Theoretical example: the model

- **PD classes**
  - 0.01 and 0.05

- **EAD segments**
  - Segment 1: $EAD = 1.0 \times OS$
  - Segment 2: $EAD = 1.5 \times OS$

- **LGD segments**
  - Segment 1: Standard LGD = 0.3
  - Segment 2: Standard LGD = 0.6
Theoretical example: PD

<table>
<thead>
<tr>
<th>PD class</th>
<th># Customers</th>
<th># Facilities</th>
<th>Exposure (EAO)</th>
<th># bad customers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>100</td>
<td>5,000</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>200</td>
<td>200</td>
<td>10,000</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>300</td>
<td>300</td>
<td>15,000</td>
<td>11</td>
</tr>
</tbody>
</table>

Exposure weighted PD = number weighted PD = observed bad rate = 3.67%

So perfect calibration!
Theoretical example: EAD/LGD distributions bads

<table>
<thead>
<tr>
<th>PD segment</th>
<th>EAD segment</th>
<th>LGD segment</th>
<th># bad customers</th>
<th>Exposure (EAO)</th>
<th>Observed EAD</th>
<th>Observed loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>50</td>
<td>75</td>
<td>45</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>100</td>
<td>100</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>400</td>
<td>600</td>
<td>360</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>11</td>
<td>550</td>
<td>775</td>
<td>435</td>
</tr>
</tbody>
</table>

Predicted EAD = observed EAD = 775
Predicted loss = observed loss = 435
Perfect calibration!
Theoretical example: what about calibration of EL?

• Depends on distribution of ‘goods’ over EAD/LGD segments!

<table>
<thead>
<tr>
<th>PD segment</th>
<th>EAD/LGD segment</th>
<th>Case 1: Underestimation</th>
<th>Case 2: Correct estimation</th>
<th>Case 3: Overestimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>20</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>79</td>
<td>84</td>
<td>80</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>138</td>
<td>35</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>52</td>
<td>155</td>
<td>185</td>
</tr>
<tr>
<td>EL</td>
<td></td>
<td>279</td>
<td>435</td>
<td>479</td>
</tr>
</tbody>
</table>

Case 1: EL is underestimated with 36%
Case 2: EL is correctly estimated
Case 3: EL is overestimated with 10%

Question: Why?
Bad rate per EAD segment differs for these three cases

Case 1: underestimation
Case 2: correct estimation
Case 3: overestimation
And PD is not calibrated on EAD segment level.

Conclusions:
• Independence assumption not true for case 1 and 3
• Can result in upward/downward bias of EL
The real case
The data

• Data from 2004 – 2009

• Data split in ‘bads’ and ‘goods’

• ‘Bads’:  
  - Take all facilities in EAD dataset (the ‘bads’)  
  - Enrich with new PD and observed loss (from LGD dataset)

• ‘Goods’:  
  - Take all facilities  
  - Enrich with new PD and apply EAD/LGD model

• Combine data of ‘goods’ and ‘bads’

• Key figures:

  # Customers > 400,000
  # current accounts > 500,000
  Total O/S > 300 million
  Total EAD > 12 million
  Total loss > 5 million
  % bads < 1%
The Basel model

- **PD model**
  - 17 dedicated risk pools

- **EAD model**
  - 6 segments

<table>
<thead>
<tr>
<th>EAD Segments</th>
<th>Compensation effect</th>
<th>Negative parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segment 1</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Segment 2</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Segment 3</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Segment 4</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Segment 5</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Segment 6</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

- **LGD model**
  - 3 segments
The loss on portfolio level is overestimated with 7%

- Overestimation smaller on whole portfolio than on ‘bads’
- Not LGD effect but PD-EAD effect
  - Overestimation in segments 2, 4 and 6.
    - Due to negative parameters set to zero
    - Highly correlated with O/S
  - Underestimation in other sub segments
    - Due to compensation effects

<table>
<thead>
<tr>
<th></th>
<th>Overestimation loss</th>
<th>Overestimation EAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bads</td>
<td>9%</td>
<td>8%</td>
</tr>
<tr>
<td>All</td>
<td>7%</td>
<td>7%</td>
</tr>
</tbody>
</table>
Bad rate varies between EAD segments: underestimation in Segment 6 and overestimation in Segment 4.
Conclusions and recommendations

- Overestimation of loss mainly due to EAD model
  - Negative parameters set to zero
  - Partially damped by PD-EAD correlation

- EL should only be analyzed at portfolio level
  - Unless calibration has been done on lower level

- EL validation should be part of model development (if possible)

- Data collection should be set up to make this possible

- Pitfall: inconsistency in EAD/LGD data collection!
Questions?