A Two Stage Mixture Model for Predicting EAD

Credit Scoring & Credit Control, Edinburgh 2013

Mindy Leow & Jonathan Crook
Credit Research Centre, University of Edinburgh, UK
Motivation

Regulatory Capital

Basel II requires each bank in a complying jurisdiction to hold

\[ \text{Regulatory Capital} \geq \text{MRCR}_{\text{credit risk}} + \text{MRCR}_{\text{market risk}} + \text{MRCR}_{\text{operational risk}} \]

where

\[ \text{MRCR}_{\text{credit risk}} = EAD \cdot \text{LGD} \cdot \Phi \left( \frac{\Phi^{-1}(PD) + \sqrt{\rho} \Phi^{-1}(0.999)}{\sqrt{1 - \rho}} \right) - PD \]

Represents 99.9\(^{th}\) percentile of expected loss distribution (VaR(99.9)).

EAD is unknown before the time of default, but known the very instant the account goes into default, so although default-time variables could be used in the modelling of LGD, they cannot be used for EAD.

Economic Capital
Common variables estimated in lieu of EAD in the literature are

Loan Equivalent Exposure (LEQ) Factor,
Credit Conversion Factor (CCF)
Exposure At Default Factor (EADF)
Dependent variables used in literature 1

Notation

\( T \) = time of default
\( B_t \) = Balance at time \( t \)
\( L_t \) = Limit at time \( t \)

\[ EAD_{t,T} = E_t(B_{t=T} \mid t \leq T, x, T - t) \]

Dependent variables used in the literature

Loan Equivalent Exposure (LEQ) Factor,

\[ EADF_{t=T} = E_t\left( \frac{B_{t=T}}{L_{t=T-l}} \mid t \leq T, x, T - t \right) \quad \text{for} \quad L_{t=T-l} \neq 0 \]

\[ EAD_{t,t=T} = L_{T-l} \ast EADF_{t=T} \]

Empirically limit usually taken at time account opened but limit likely to have changed since.
Dependent variables used in literature 2

Credit Conversion Factor (CCF)

\[ CCF_{t=T} = \begin{cases} 
E_t \left( \frac{B_{t=T}}{B_{t=T-l}} \right) | t \leq T, x, T - t 
& \text{if } B_{t=T-l} \neq 0 \\
0 
& \text{if } B_{t=T-l} = 0 
\end{cases} \]

\[ EAD_{t,T} = B_{t=T-l} \ast CCF_{t=T} \]

Tries to get better est of balance at default by taking account of balance at some obsn. point before default. But balance at time of observation could be 0 or negative

Exposure At Default Factor (EADF)

\[ LEQ_{t=T} = \begin{cases} 
E_t \left( \frac{B_{t=T} - B_{t=T-l}}{L_{t=T-l} - B_{t=T-l}} \right) | t \leq T, x, T - t 
& \text{if } L_{t=T-l} \neq B_{t=T-l} \\
0 
& \text{if } L_{t=T-l} = B_{t=T-l} 
\end{cases} \]

\[ B_{t=T} = B_{t=T-l} + LEQ_{t=T} (L_{t=T-l} - B_{t=T-l}) \]

- Not defined when \( L_{t=T-l} = B_{t=T-l} \)
- Positive \( LEQ_T \) could be due to different situations with different characteristics
  Negative \( LEQ_T \) likewise
The Literature

EAD papers in the corporate sector

• Araten and Jacobs (2001)
• Jimenez et al. (2009)
• Jacobs (2008)
• Jimenez and Mencia (2009)

Retail Loans

• Qi (2009)
Distribution of Balances relative to credit limit at default

Extent of balances with reference to credit limit, at time of default
for bal/lim < 3

PERCENT

ratio of balance to limit
Proposed new methodology

Two-step mixture model

Let $t =$ duration time; $T =$ event time; $B =$ balance, $L =$ limit

Wish to predict, at time $t = 0$, outstanding balance at time of default ($t = T$)

Estimate Survival model to predict, at $t = 0$: $P(B_{it} > L_{it})$

Estimate model to predict, at $t=0$: $L_{it}$

Estimate model to predict, at $t=0$: $B_{it}$

We wish

$$E_{t=0}(B_{it}) = \{P(B_{it} \geq L_{it}) \times E_{t=0}(L_{it} \mid B_{it} \geq L_{it})\} + \{P(B_{it} < L_{it}) \times E_{t=0}(B_{it} \mid B_{it} < L_{it})\}$$
Data

Credit cards opened between 2001 and 2010

Minimum payment computed as percentage of balance end of previous month

Account goes into default state if is 3 months in arrears (not necessarily consecutive months)

Accounts removed:

• on books less than 9 months
• those with zero credit limit at any time
Portfolio of loans

Defaults

Training set:
accounts opened up to end 2008
* used to estimate
\( P(B \geq L) \)

Limit training set:
Accounts that have balance \( \geq \) limit at any point in the loan
* used to estimate \( L \)

Balance training set accounts where balance < limit throughout the loan
* used to estimate \( B \)

Non-Defaults

Test set I:
accounts opened from beginning 2009

Test set II
accounts opened from beginning 2009 that are in default

Predicted balance =
\[
\{ P(B \geq L) \times \hat{L} \} \\
+ \\
\{ (1 - P(B \geq L)) \times \hat{B} \}
\]
Methodology 2: Survival model

\[ S_{it} = \begin{cases} 1 & \text{if } B_{t=T} \geq L_{it=T} \\ 0 & \text{otherwise} \end{cases} \]

\[
\log \left( \frac{S_{it}}{1 - S_{it}} \right) = \nu + \beta_1 X_i + \beta_2 Y_{i,t-6} + \beta_3 Z_{t-6}
\]

Estimated using discrete time, \textit{repeated events} survival estimators
Methodology 3: conditional balance and limit equations

General specification

Panel estimators with account specific random effects (SEs adjusted for first order serial correln.)

\[ y_{it} = \mu + \gamma_1 X_i + \gamma_2 Y_{it-6} + \gamma_3 Z_{t-6} + \alpha_i + \varepsilon_{it} \]

Assumptions include
\[ \alpha_i \sim IID(0, \sigma_\alpha^2) \quad \varepsilon_{it} \sim IID(0, \sigma_\varepsilon^2) \]

Samples:
conditional Limit: accounts where \( B_{it} \geq L_{it} \) for any \( t \) gives \( E_{t=0}(L_{it} \mid B_{it} \geq L_{it}) \)
conditional Balance: accounts where \( B_{it} < L_{it} \) for any \( t \) gives \( E_{t=0}(B_{it} \mid B_{it} < L_{it}) \)
Methodology 4: unconditional balance equation

\[
\tilde{B}_{it} = \left\{ P(B_{it} \geq L_{it}) \times \hat{L}_{it} \right\} + \left\{ (1 - P(B_{it} \geq L_{it})) \times \hat{B}_{it} \right\}
\]

- Predicted balance account \(i\) time \(t\)
- From survival model
- From limit (panel) model
- From balance (panel) model
### Parameter estimates

<table>
<thead>
<tr>
<th>Application Variables</th>
<th>Survival</th>
<th>Balance</th>
<th>Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at appl (10 groups)</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Employment status (5 groups)</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Income, log</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Binary ind for 0 or missing income</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Landline (Y/N)</td>
<td>Not sig</td>
<td>+</td>
<td>Not selected</td>
</tr>
<tr>
<td>No of cards</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Time at address</td>
<td>Not sig</td>
<td>Not selected</td>
<td>Not selected</td>
</tr>
<tr>
<td>Time with bank</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Binary ind., missing unk TwB</td>
<td>-</td>
<td>Not selected</td>
<td>Not selected</td>
</tr>
<tr>
<td>X (5 groups)</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Behavioural Variables lagged 6 months</th>
<th>Survival</th>
<th>Balance</th>
<th>Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average trans value</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>No cash withdrawals</td>
<td>+</td>
<td>Not selected</td>
<td>Not selected</td>
</tr>
<tr>
<td>Amount of cash withdrawal</td>
<td>Not sig</td>
<td>Not selected</td>
<td>Not selected</td>
</tr>
<tr>
<td>Credit Limit</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Rate of total jumps</td>
<td>+</td>
<td>+</td>
<td>Not selected</td>
</tr>
<tr>
<td>% of months in arrears</td>
<td>+</td>
<td>-</td>
<td>Not selected</td>
</tr>
<tr>
<td>Repayment amount</td>
<td>Not selected</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Outstanding balance</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Macroeconomic variables lagged 6 months</th>
<th>Survival</th>
<th>Balance</th>
<th>Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average wage earnings</td>
<td>Not sig</td>
<td>Not selected</td>
<td>Not selected</td>
</tr>
<tr>
<td>Credit card interest rate</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Consumer confidence</td>
<td>+</td>
<td>Not selected</td>
<td>Not selected</td>
</tr>
<tr>
<td>House price index</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Index of production</td>
<td>-</td>
<td>Not selected</td>
<td>Not selected</td>
</tr>
<tr>
<td>Base interest rate</td>
<td>-</td>
<td>Not selected</td>
<td>Not selected</td>
</tr>
<tr>
<td>Amount outstanding, Ln</td>
<td>+</td>
<td>Not selected</td>
<td>-</td>
</tr>
<tr>
<td>FTSE. Ln</td>
<td>-</td>
<td>-</td>
<td>Not selected</td>
</tr>
<tr>
<td>RPI</td>
<td>Not sig</td>
<td>+</td>
<td>Not selected</td>
</tr>
<tr>
<td>Unemployment index</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model Specific Variables</th>
<th>Survival</th>
<th>Balance</th>
<th>Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survival time to next event</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No times event has happened</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time on books</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.250</td>
<td>0.250</td>
<td></td>
</tr>
<tr>
<td>N*T</td>
<td>&gt;800k</td>
<td>&gt;900k</td>
<td></td>
</tr>
</tbody>
</table>
Results 1: Predictive performance: test sets

Compare observed and predicted $B_{it}$

<table>
<thead>
<tr>
<th>Model</th>
<th>Test set</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixture</td>
<td>1: defaults, all observations</td>
<td>0.4989</td>
</tr>
<tr>
<td>Mixture</td>
<td>2: defaults, default time only</td>
<td>0.5517</td>
</tr>
</tbody>
</table>
Results 2: Observed and Predicted Distributions (all observed accts)

Comparative histogram of observed and predicted balances, for balance less than 9600, indexed 'TES, LOT9, PROB(repSURV), BAL(RE), LIM(RE)'

Observed

Predicted
Results 3: Observed and Predicted Distributions (at default time, default accounts only)

Comparative histogram of observed and predicted balances, for balance less than 9600, indexed 'TES, LOT9, PROB(repSURV), BAL(RE), LIM(RE), AT DEFAULT TIME OBSERVATIONS ONLY'
Use of a two stage mixture model to predict the amount outstanding, six months ahead for individual credit card accounts is feasible and gives reasonably accurate results.

Future work:

We have now included macroeconomic variables (MEVs) in predictive models of PD, LGD and EAD. We plan to refine these are use them to run stress tests for RWA using MC simulation from observed historical distributions of MEVs.

To try to understand the relationships between behavioural variables and macroeconomic variables.