UNIVERSITY OF EDINBURGH
Business School

CRC Credit Research Centre

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

UNIVERSITY OF EDINBURGH
Business School

CRC Credit Research Centre

Motivation

Regulatory Capital

Basel II requires each bank in a complying jurisdiction to hold

Regulatory Capital ≥ $MRCR_{\text{credit risk}} + MRCR_{\text{market risk}} + MRCR_{\text{operational risk}}$

where $MRCR_{\text{credit risk}}$

$$= EAD.LGD * \left(\Phi\left[\frac{\Phi^{-1}(PD) + \sqrt{\rho}\,\Phi^{-1}(0.999)}{\sqrt{1-\rho}} \right] - PD \right)$$

Represents 99.9th%ile 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

Alternative dependent variables

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

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

Dependent variables used in the literature

Loan Equivalent Exposure (LEQ) Factor,

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

$$\mathrm{EAD}_{t,t=T} = L_{T-l} * EADF_{t=T}$$

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)

$$CCF_{t=T} = \begin{cases} E_t\left(\frac{B_{t=T}}{B_{t=T-l}} \mid t \le T, \mathbf{x}, T-t\right) & \text{if} \quad B_{t=T-l} \ne 0 \\ 0 & \text{if} \quad B_{t=T-l} = 0 \end{cases}$$

$$EAD_{t,T} = B_{t=T-l} * CCF_{t=T}$$

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)

$$LEQ_{t=T} = \begin{cases} E_t\left(\frac{B_{t=T} - B_{t=T-l}}{L_{t=T-l} - B_{t=T-l}} \mid \mathrm{t} \le \mathrm{T}, \mathbf{x}, T-t\right) & \text{if} \quad L_{t=T-l} \ne B_{t=T-l} \\ 0 & \text{if} \quad L_{t=T-l} = B_{t=T-l} \end{cases}$$

$$B_{t=T} = B_{t=T-l} + LEQ_{t=T}(L_{t=T-l} - B_{t=T-l})$$

- 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

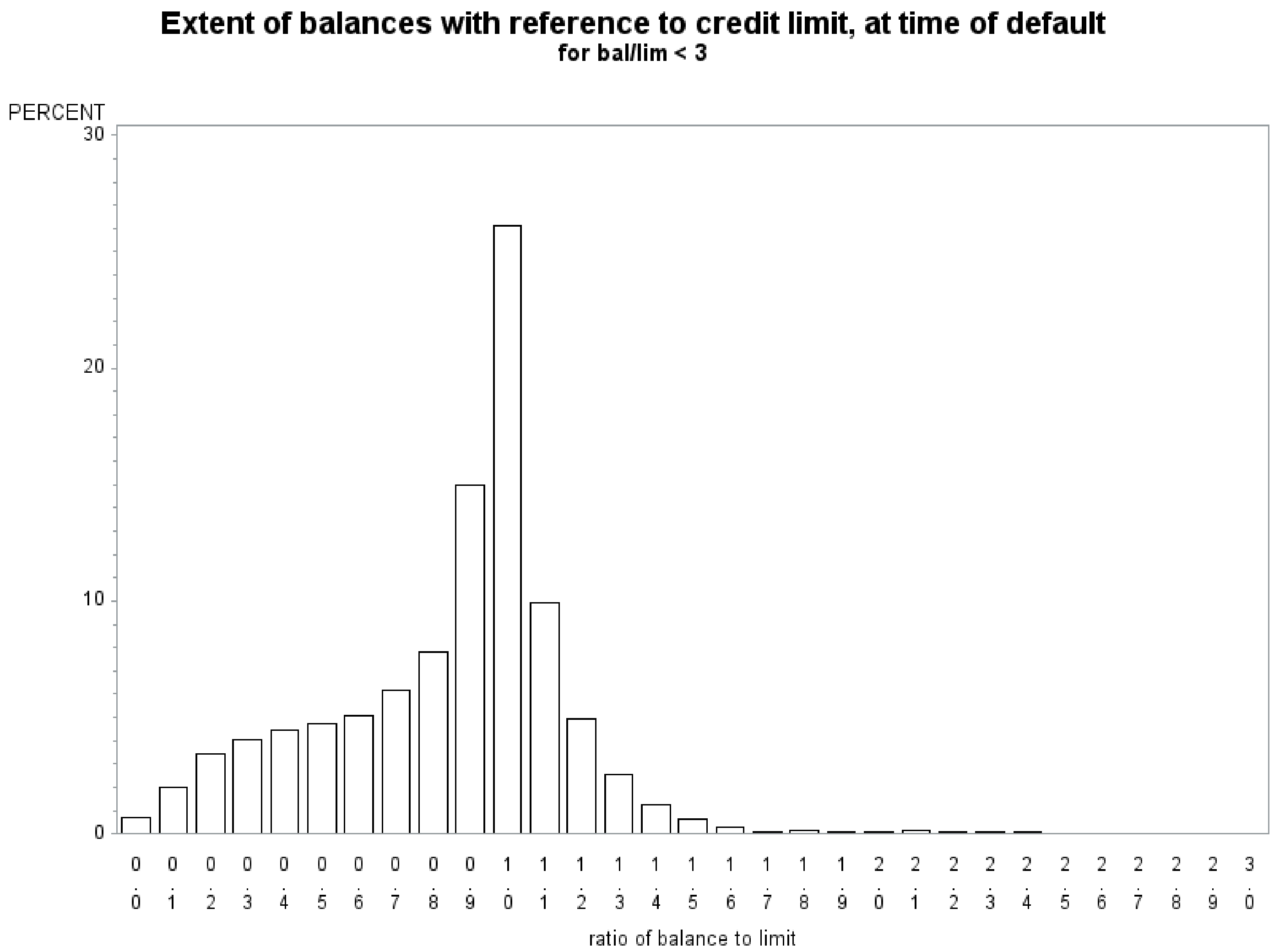
EAD papers in the corporate sector

- 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



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)

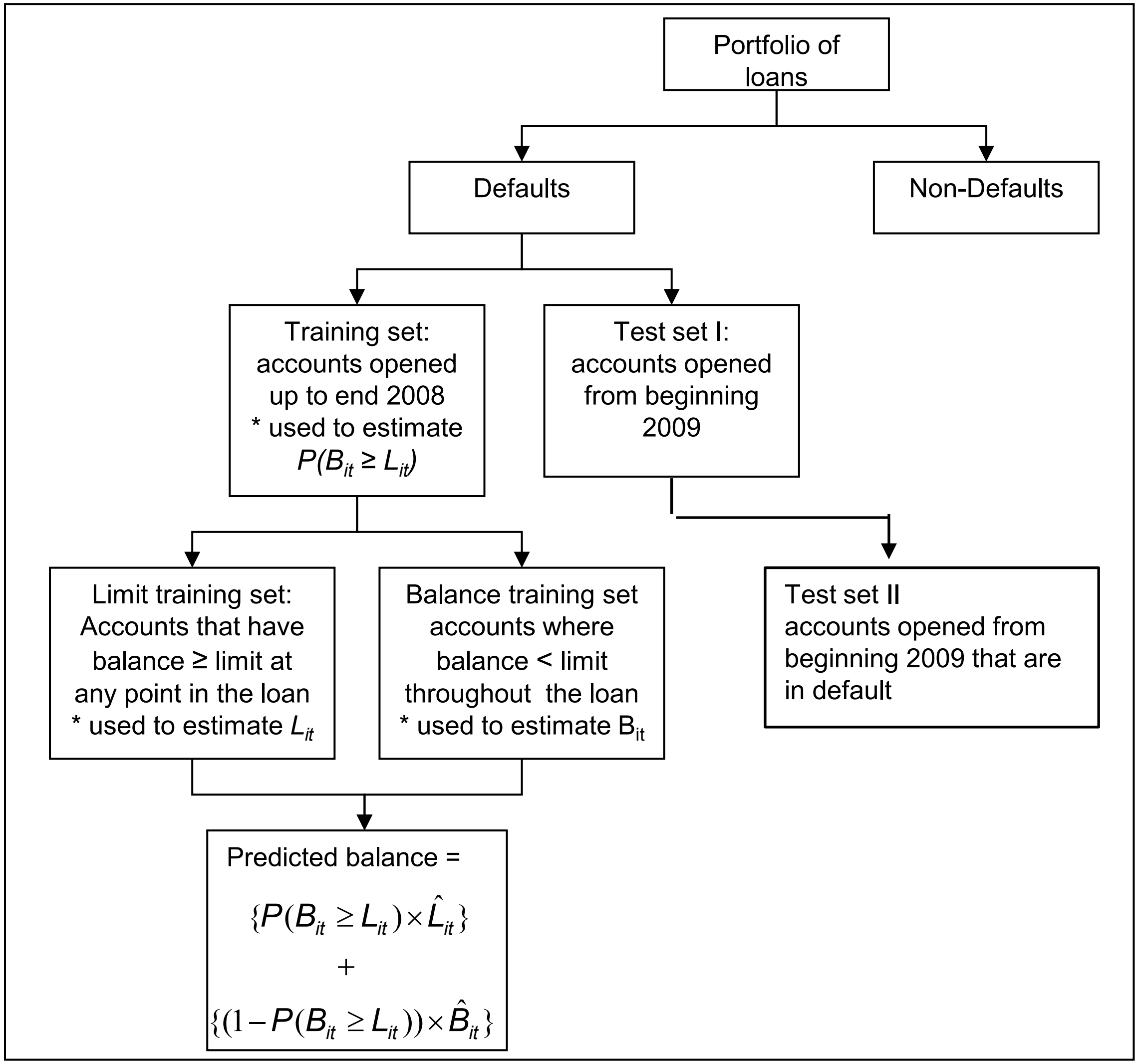
Accounts removed:

- on books less than 9 months
- those with zero credit limit at any time

UNIVERSITY OF EDINBURGH Business School

CRC Credit Research Centre

Flow chart of methodology



Methodology 2: Survival model

$$S_{it} = \begin{cases} 1 & \text{if} \quad B_{t=T} \geq L_{i\,t=T} \\ 0 & \text{otherwise} \end{cases}$$

$$\log\left(\frac{S_{it}}{1-S_{it}}\right) = \nu + \beta_1 \mathbf{X}_i + \beta_2 \mathbf{Y}_{i,t-6} + \beta_3 \mathbf{Z}_{t-6}$$

- $\mathbf{X}_i$: account dependent, time independent (application vars)
- $\mathbf{Y}_{i,t-6}$: account dep, time dep (behavl. vars)
- $\mathbf{Z}_{t-6}$: account indep, time dep (macroeonomic vars)

Estimated using discrete time, *repeated events* survival estimators

UNIVERSITY OF EDINBURGH Business School

CRC Credit Research Centre

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 $\alpha_i \sim IID(0, \sigma_\alpha^{\ 2})$ $\varepsilon_{it} \sim IID(0, \sigma_\varepsilon^{\ 2})$

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

From survival model

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

Predicted balance account *i* time *t*

From limit (panel) model

From balance (panel) model

Parameter estimates

	Survival P(B>L)	Balance	Limit
Application Variables			
Age at appl (10 groups)			
Employment status (5 groups)			
Income, log	-	+	+
Binary ind for 0 or missing income	-	+	+
Landline (Y/N)	Not sig	+	Not selected
No of cards	-	+	+
Time at address	Not sig	Not selected	Not selected
Time with bank	-	+	+
Binary ind. ,missing unk TwB	-	Not selected	Not selected
X (5 groups)	+	-	-
Behavioural Variables lagged 6 months			
Average trans value	-	+	+
No cash withdrawals	+	Not selected	Not selected
Amount of cash withdrawal	Not sig	Not selected	Not selected
Credit Limit		+	+
Rate of total jumps	+	+	Not selected
% of months in arrears	+	-	Not selected
Repayment amount		Not selected	+
Outstanding balance		+	+

	Survival	Balance	Limit
Macroeconomic variables lagged 6 months			
Average wage earngs	Not sig	Not selected	Not selected
Credit card interest rate	+	+	-
Consumer confidence		+	Not selected
House price index	-	-	+
Index of production	-	Not selected	Not selected
Base interest rate	-	Not selected	Not selected
Amount outstanding, ln	+	Not selected	-
FTSE. Ln	-	-	Not selected
RPI	Not sig	+	Not selected
Unemployment index		+	-
Model Specific Variables			
Survival time to next event	-		
No times event has happened	+		
Time on books		+	+
ρ		0.250	0.250
N*T		>800k	>900k

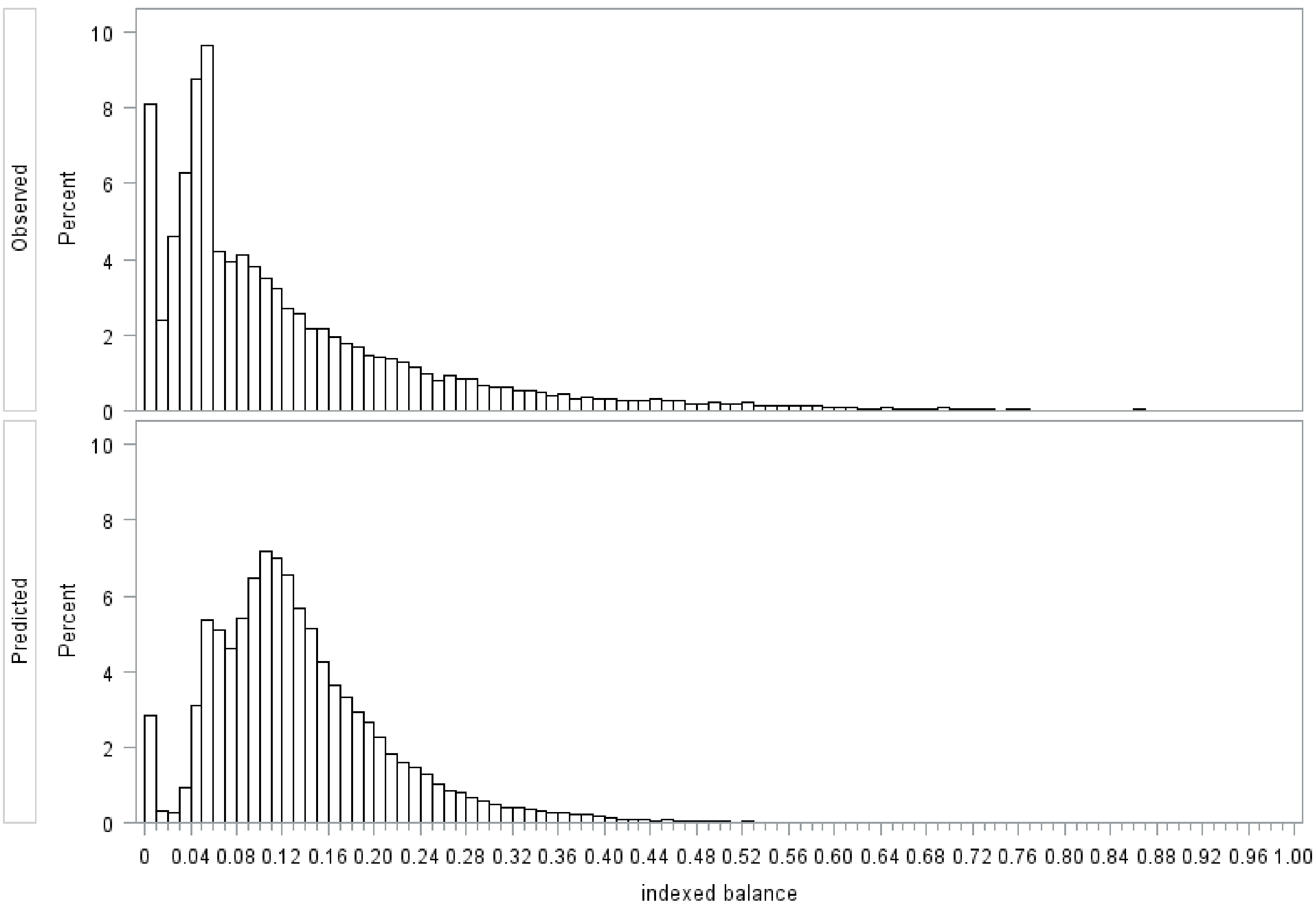
Results 1: Predictive performance: test sets

Compare observed and predicted B_{it}

Model	Test set	R^2
Mixture	1: defaults, all observations	0.4989
Mixture	2: defaults, default time only	0.5517

Results 2: Observed and Predicted Distributions (all observed accts)

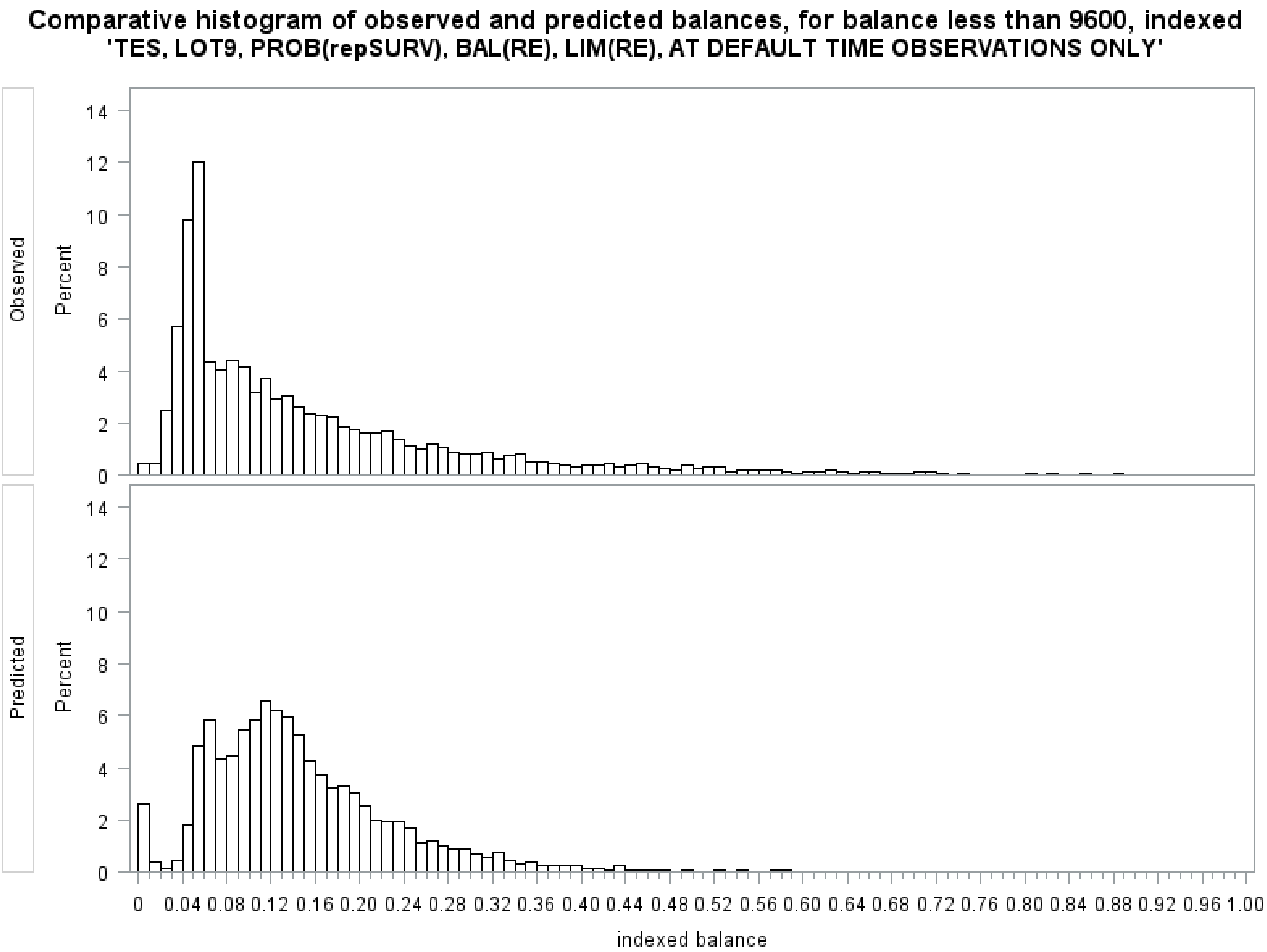
Comparative histogram of observed and predicted balances, for balance less than 9600, indexed 'TES, LOT9, PROB(repSURV), BAL(RE), LIM(RE)'



Observed

Predicted

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



Observed

Predicted

Conclusions

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.