Dynamic Consumer Risk Models: An Overview

Jonathan Crook and Tony Bellotti

Credit Research Centre
Management School and Economics
University of Edinburgh

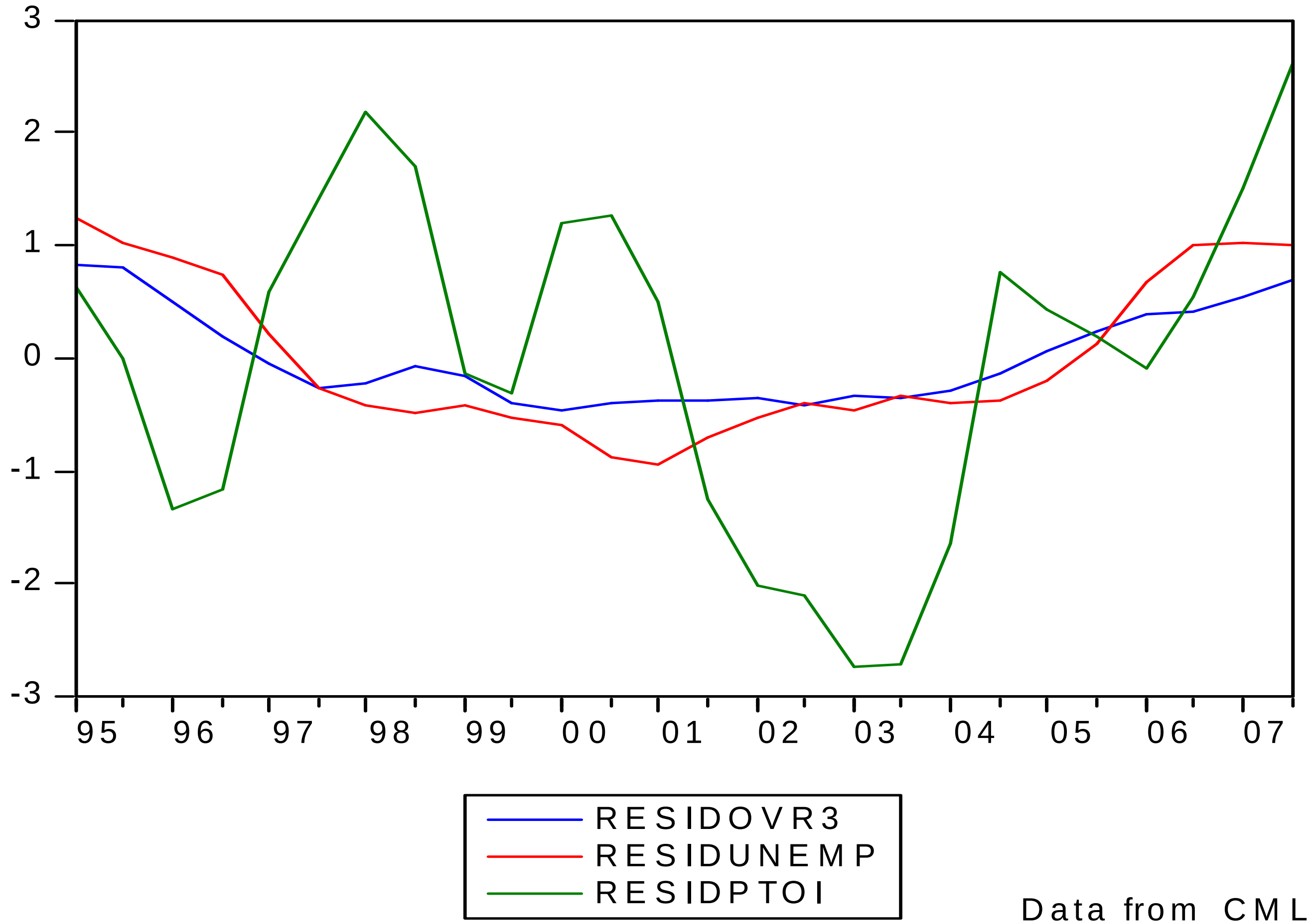
www.crc.man.ed.ac.uk

Research funded by EPSRC grant EP/D505380/1, working as part of the **Quantitative Financial Risk Management Centre**.

Structure

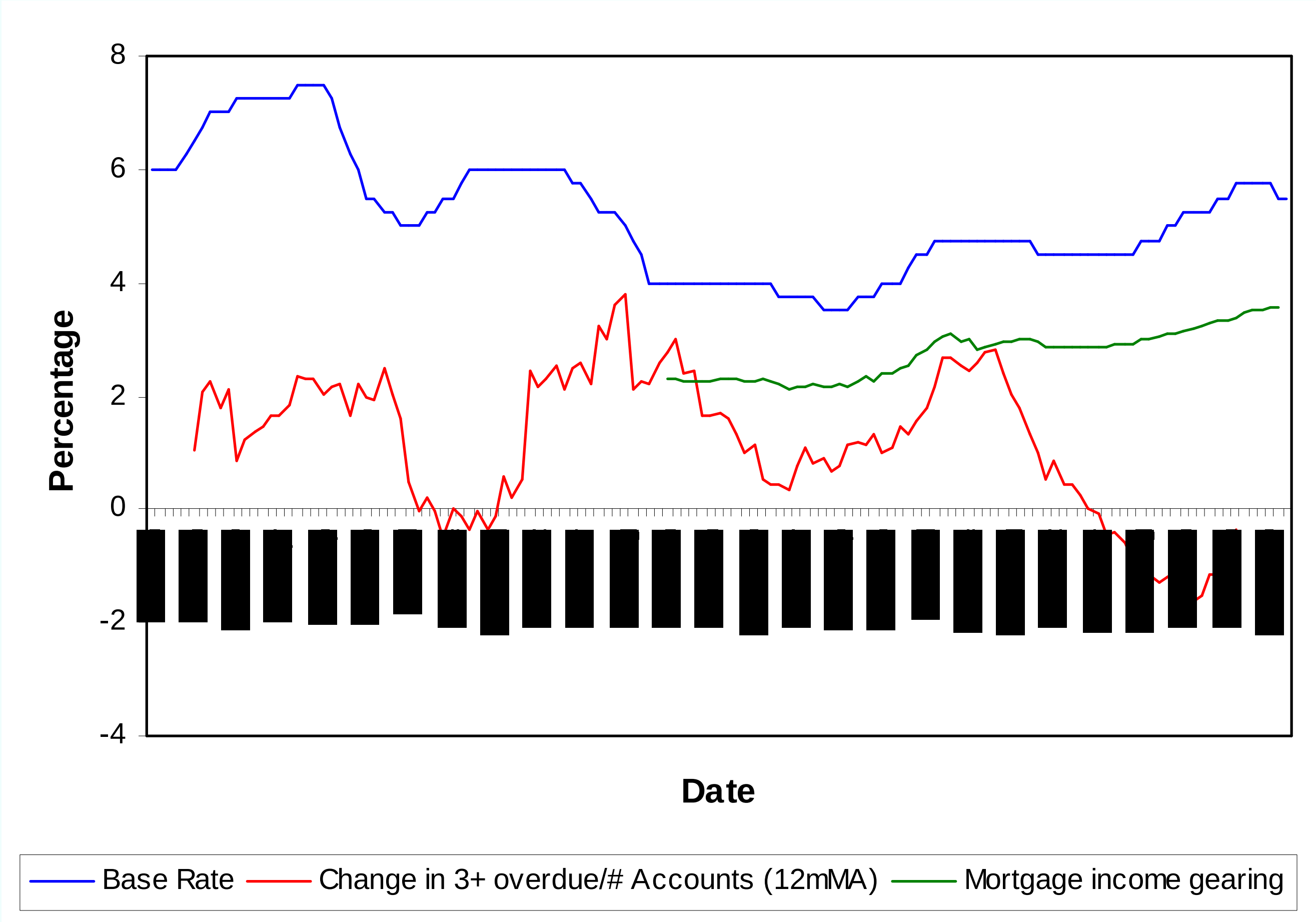
- Dynamic Modelling at the level of the account
- Dynamic Modelling for Portfolios

Mortgage Delinquency and State of the Economy (detrended)



Data from CML & ONS

UK Credit Cards Change in 3 + Overdue as % of no of Accounts



Data from APACS & ONS

Aggregate Consumer Default Rates and the Economy (time series)

- US: Sullivan (1987)
  Ausubel (1997)
  Grieb (2001)
  Banasik & Crook (2005)

- UK: Whitely (2004)

This presentation is about cross-section –time series models

Generic Model

$$PD_{it} = P(d_{it} = 1) = f\left(\sum_{m=0}^{M} \beta_m x_{mi} + \sum_{l=0}^{L} \sum_{p=0}^{P} \beta_{pl} x_{pit-l} + \sum_{l}^{L} \sum_{j=0}^{J} \gamma_{kl} Z_{jt-l} + \text{interactions} + \varepsilon_{it} \right)$$

Application Model

Behavioural Model

Why do borrowers default?

- Strategic (fraud or decline in asset values)
- Unexpected net income shock (unemployment, health costs, divorce, children etc)

We try to predict if a borrower will strategically default or unexpectedly suffer unemployment, health costs, divorce etc using

x_i Variables that we do not observe to change

x_{it-l} Variables that we do observe to change over time

Modellers Should include exogenous variables that change over time eg are you married in t-1,t-2, income in t-1, t-2 etc

Instead modellers actually use

- endogenous variables e.g. repayment performance in t-1, t-2 etc
- exogenous variables measured only at one point in time e.g. application variables

Data Format

- Equation (1) relates to a (possibly unbalanced) panel.

d_{it} (possibly) taking on 0 or 1 until default then missing

x_i Staying constant over *t* for any *i*

x_{it-l} Varying over *t* for any *i* (missing for written off cases after writeoff)

Z_t Constant across *i*, varying over *t*

- Lenders have panel format data.

Survival Models

- Interested in probability at an instant in time, of leaving one state e.g. “being up to date” and entering another state “90 days overdue”.

- Probability of default (repaying early) in next instant of time, conditional on not having defaulted (repaid early) before is

$$\lambda(t) = \underset{\Delta t \to 0}{Lim} \frac{P(t \le T < t + \Delta t \mid T \ge t)}{\Delta t}$$

Cox Proportional hazards commonly used.

$$\lambda_i(t, x_t) = \lambda_0(t) \exp(\mathbf{x_i^T \beta})$$

Dynamic only in sense that there is a common baseline hazard that varies with time (and is shifted according to covariates).

Introducing Macroeconomic Variables into Survival Models

- If use "Cox's PH"

$$\lambda_i(t, x(t), \beta) = \lambda_0(t)\exp(\mathbf{x}_i^T(\mathbf{t})\boldsymbol{\beta})$$

- Bellotti & Crook (2007)
  - Included macroeconomic variables and interactions
  - If include MEVs predictive performance increases
  - Interest rates, real earnings, consumer confidence
  - Need to predict future values of MEVs
    (But if include lagged MEVs predictive performance also increased

Advantages of Survival Models over Static Logistic Regression

- Can predict probability of default in next time period, given has not defaulted before, not just some time in a predefined time period (e.g.12 months)
- Likelihood function takes into account censored observations
- Can use it to predict probability of surviving in each time period so can use to predict profitability

Panel Techniques

- Time periods treated as discrete

- Can predict probability of events that may reoccur
  - Example: probability borrower will miss one payment in period t

- Can predict one off events e.g. default (Cox's discrete hazard function)
  - Example Saurina & Traucarte (2007)
    - Yearly time periods. 2.94 million mortgages in Spain
    - Predicted probability borrower will miss third payment in a particular year.
    - Covariates include whether had defaulted in past and GDP
    - AUROC 0.78

  - Example Valles (2006)
    - Yearly time periods, corporate defaults
    - Predicted probability 90 days overdue
    - Covariates includedGDP growth (-) inflation rate (-), unemployment (+)

Simplistic Example of Panel Estimation

- Probability of missing one payment in month t
- Sample of credit cards issued late 90s –early 2000s
- Random effects probit

$$d_{it}^{*} = \mu + \mathbf{x}_{\mathbf{i}}^{\mathbf{T}}\boldsymbol{\beta} + \mathbf{x}_{\mathbf{it}}^{\mathbf{T}}\boldsymbol{\gamma} + \mathbf{z}_{\mathbf{t}}^{\mathbf{T}}\boldsymbol{\delta} + \alpha_i + \varepsilon_{it}$$

- Macroeconomic variables (time varying)
- Duration time
- 50% (approx) of variance in errors due to random effect

Application Model (Selected Parameters)

Variable	Sign	z
Macroeconomic Variables		
interest rate	+	18.8*
house price index	-	-22.6*
unemployment index	+	2.1
consumer confidence	-	-0.1
Application Variables		
income	+	0.6
age 25-29	-	-6.1*
age 30-33	-	-5.8*
age 34-37	-	-5.9*
age 38-41	-	-5.7*
age 42-47	-	-7.5*
age 48-55	-	-9.8*
age 56+	-	-9.0*
Duration time	+	+35.7*
Duration time sqrd	-	-32.7*

Wald (Chi sqrd)	3969*	
Rho Variance random effect/ Total variance	0.49*	

* denotes significance @1%

Behavioural Model (selected parameters)

Variable	Sign	z
Macroeconomic Variables		
interest rate	+	8.6*
house price index	-	-1.1
unemployment index	+	10.9*
consumer confidence	+	0.4
Application Variables		
income	+	2.2
age 25-29	-	-3.4*
age 30-33	-	-3.4*
age 34-37	-	-3.7*
age 38-41	-	-4.4*
age 42-47	-	-6.3*
age 48-55	-	-8.2*
age 56+	-	-7.5*
Duration time	+	32.2*
Duration time sqrd	-	21.2*
Behavioural Variables (lagged 1 month)		
x1	-	14.1*
x2	+	21.5*
x3	+	1.6
balance/credit limit	+	27.9*
x1 sqrd	+	21.4*
x2 sqrd	-	-22.3*
x3 sqrd	+	0.7
balance/credit limit sqrd	+	10.4*
Wald (Chi sqrd)	8687*	
Rho Variance random effect/ Total variance	0.45*	

* denotes significance @1%

Correction Factor Models

- De Andrade (SMEs)
- Use LR to predict default

$$PD_{it} = \frac{1}{1+\exp(-\hat{f}_{it} - cf_t)}$$

- cf_{it} is (observed default rate in time t /predicted default rate in time t)
- Observed/predicted is predicted from a time series model (ADL) explaining this in terms of macroeconomic variables
- But does not allow PD ranking of applicants to change with the economy.

Portfolio Models

- We are interested in the distribution of default rates (or losses) esp the α- precentile of the fraction of loans that default (VaR)

- Factor Models
  - Random Effects
  - Kalman Filter

- Non-Factor Models

- Reduced Form Models

Factor Models

- Proposed by Vasicek (1997), Finger (1999), Schonbucher (2000)
- First application to retail loan portfolios by Perli & Nayda (2004) [PN]
- Default probabilities may be correlated e.g. employed in same industry or subject to same interest rate changes
- PN assume consumer defaults when $V_i < K_i$

Assume $$V_{it} = \sqrt{\rho} Z_t + \sqrt{1-\rho}\, \varepsilon_{it}$$

ρ = correlation coefficient between default probabilities

ε_{it} = borrower specific noise

Z_t = common factor (observables – MEVs and/or unobservables)

Corr (V_{it} , V_{jt}) = ρ Corr(V_{it}, Z_t) = $\sqrt{\rho}$

Distribution Of The Fraction Of Borrowers That Default

$$V_{it} = \sqrt{\rho} Z_t + \sqrt{1-\rho}\, \varepsilon_{it}$$

ε_{it} ~NID(0,1)

$$V_{it} < K_0 + \mathbf{K}^{\mathbf{T}} \mathbf{Y_t} \Rightarrow \sqrt{\rho} Z_t + \sqrt{1-\rho}\, \varepsilon_{it} < K_0 + \mathbf{K}^{\mathbf{T}} \mathbf{Y_t}$$

$$\Rightarrow \varepsilon_{it} < \frac{K_0 + \mathbf{K}^{\mathbf{T}} \mathbf{Y_t} - \sqrt{\rho} Z_t}{\sqrt{1-\rho}}$$

$$\Rightarrow P(V_{it} < K_0 + \mathbf{K}^{\mathbf{T}} \mathbf{Y_t} \mid Z_t = z_t) = \Phi\left(\frac{K_0 + \mathbf{K}^{\mathbf{T}} \mathbf{Y_t} - \sqrt{\rho} z_t}{\sqrt{1-\rho}} \right)$$

If X is proportion of borrowers that actually default

$$\Rightarrow P(X \le x) = \Phi\left(\frac{1}{\sqrt{\rho}} \left(\sqrt{1-\rho}\, \Phi^{-1}(x) - K_0 - \mathbf{K}^{\mathbf{T}} \mathbf{Y_t} \right) \right)$$

From which the density function can be gained

Estimation by Random Effects

- Suppose we observe $\{D_{it} = d_{1t} ......D_{N_t} = d_{N_t}\}$ where d_{it} = 1 if borrower I defaults, 0 otherwise

- Then $$P(d_{1t}, ...d_{N_t t} \mid z_t) = \prod_{i=1}^{N_t} [P_{it}(z_t)^{d_{it}}][1 - P_{it}(z_t)]^{1-d_{it}}$$

- Integrate over z_t and sum over t:

$$L = \sum_{t=1}^{m} \ln \left\{ \int_{-\infty}^{\infty} \prod_{i=1}^{N_t} [P_{it}(z_t)^{d_{it}}][1 - P_{it}(z_t)]^{1-d_{it}} \right\} \phi(z_t) dz_t$$

Which is a random effects (with respect to time (!)) probit model.

Some Results

- Hamerle & Rosch (2006)

- Yearly data 1991-2000, 53,000 firms
- Found: asset correlation only 0.0004, not significant
- Found VaR for CM, CPV, CR+ virtually identical for several percentiles

- Rosch & Scheule (2004)

- Predicted charge-off rates. Data for 1991-2001.
- Found correlation coefficient: credit cards: 0.012, real estate loans 0.0098, other consumer loans 0.0073 – all below Basel II
- Effects of macroeconomic vars on charge-off rates for credit cards:

Credit cards	Real Estate
ΔConsumer Price Index (-)	ΔIndustrial production (-)
Δ GDP (-)	

Expected loss closer to actual when macroeconomic variables included

Kalman Filter Methods

- Can be used to relate default activity to unobserved factors.

- Method $\mathbf{y_t} = \mathbf{A_t x} + \boldsymbol{\upsilon}$ Observation equation

$$\mathbf{x_t} = \boldsymbol{\theta}\mathbf{x_{t-1}} + \boldsymbol{\omega}_\mathbf{t}$$ State equation

Use $$\mathbf{x_t^{[t-1]}} = \boldsymbol{\theta}\mathbf{x_{t-1}^{[t-1]}}$$

$$\mathbf{x_t^{[t]}} = \mathbf{x_t^{[t-1]}} + \mathbf{k_t}(\mathbf{y_t} - \mathbf{A_t x_t^{[t-1]}})$$

To estimate $\boldsymbol{\theta}, \mathbf{k}_t \mathbf{A_t}$

Example

- Jimenez & Mencia (2007)

- Vector Auto regression (VAR) to explain
  Growth in number of loans & Growth in default frequency
- Quarterly data from Credit Register of Spain 1984-2006 10 sector plus consumer loans and mortgages

Findings

- Δdefault rates related to (lagged) GDP (-)
  latent factor (+)
  but not interest rates
- Took random values of macroeconomic variables, errors and latent factors to simulate Loss on portfolio.
- When latent factors included VaR (99.9%) was 5% (consumer loans) 2% (mortgages) lower than when latent factors left out

Non-Factor Models

- Rodriguez & Trucharte (2007)

- Pooled panel estimates to predict PD. Classify into risk classes to find distribution in each year. Take randm samples of loans, allocate to reisk classes until have first distribution.

- Pooled simulated loans over all years (1990-2004). Find loss rates (as % of exposure) were higher than rates covered by Basel.

- For 2004 they stress the predictors.

  Found loss rates at all percentiles, for the worst year in the data period were much larger than implied by Basel using an average PD over the cycle. E.g 99%ile: Basel loss rate 6.4%, worst case scenario loss rate 11.4%

Reduced Form Models

Cause of default does not depend on asset values

- Markov Chains
- Stochastic Intensity Models

Markov Chains

- Panel data can be expressed as transition matrix

To fix ideas

Delinquency states

	1	2	3	V
1	$p_{11}(t,t+1)$	$p_{12}t,t+1)$	$p_{13}(t,t+1)$	$p_{1V}(t,t+1)$
2	$p_{21}(t,t+1)$			
3				
.				
.				
V	0	0	0	0

(Time) homogeneous: if p_{uv} does not depend on time

First order: if $p(X_t=u_t)$ depends only on state in t-1

Wth order

$$P^i(X_t = u_t \mid X_{t-1} = u_{t-1}, \ldots X_1 = u_1) = P^i(X_t = u_t \mid X_t = u_t, \ldots\ldots X_{t-w} = u_{t-w})$$

Some Observations

- Can model p_{uv} as logit

$$\ln\left(\frac{p_{uv}}{1-p_{uv}}\right)_i = \mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta} + \mathbf{Z}_\mathbf{t}^\mathbf{T}\boldsymbol{\gamma}$$

- If γ is significant, the MC is not (time) homogeneous
- If logit includes (significant) lags up to *t-l* of dependent variable then MC is order *t-l*

- If use discrete survival model we are modelling (X_t = state X time t, 1 = default 0 = not default)

$$P(X_t = 1 \mid X_{t-1} = 0, X_{t-2} = 0.....X_0 = 0)$$

Published Findings

- Transition probabilities not first order (and probably not second order either) Ho et al (2004), Till and Hand (2001)
- Mover-Stayer too simplistic. Possibly those who stay; move up to 3 times, those that move 4 times, those that move 5+ times in 48 months (Ho et al)
- First hitting times (to 3 overdue): from 0, 1, 2 payments overdue is 124, 108, 69 months (Till & Hand)

More Recent Applications

Embed Mcs into Markov Decision Processes

Example: Trench (2003)

State = combination of (a) management control variables e.g. APR, credit limit (b) customer behaviour variables

Choose action to max NPV of cash flows using Mc (assumed first order) to solve dynamic optimisation.

$$V_t(s) = \underset{a \in A_S}{Max}\{NCF(s_a) + \beta \sum_{v \in S} P(v \mid s_a) V_{t+1}(v)\}$$

Stochastic Intensity Models

- Assume continuous time.
- Assume a Poisson process, value N_t (assumed integer) at *t.* Probability of increase in *N* by 1 state in *dt* is λdt
- Regard change in *N* as jump from state *u* to default. Can form generator matrix

$$\begin{pmatrix} -\lambda_{11} & \lambda_{12} & \dots\dots\dots & \lambda_{1V} \\ \lambda_{21} & \lambda_{22} & & \\ . & & & \\ 0 & \dots & & 0 \end{pmatrix}$$

λ_{uv} = prob chain in state *v* at time *t*, given was in state *u* at time 0

Can be modelled as hazard rates with time varying covariates (Macroeconomic variables)