Hierarchical Bayesian modelling of exposure at default for revolving credit facilities

Mark Bass

Credit Scoring & Credit Control XVI

August 28-30, 2019

Outline

1. Some background on the problem.
2. The data used for model fitting/testing. We use Barclays' view of GCD data.
3. Model specifications. Reparameterisation, sampling from posterior and predictive distributions.
4. Results. Model selection. Comparison with Tobit model.
5. Discussion and future work.

Introduction

- Exposure at default (EAD) is an estimate of the exposure against a facility upon default of an obligor.
- EAD feeds into the calculation of RWA.
- CRR refers to credit conversion factors (CCF), the proportion of available headroom drawn between observation and default.
- By definition we have $EAD = B_0 + CCF(L_0 - B_0)$, where B_0 and L_0 are the balance and limit at observation.
- Banks with advanced permissions can model their own CCF (CRR Article 166).
- The time horizon for EAD estimates is one year (SS11/13 14.11).
- Estimates for EAD must be no less than current drawings, hence $CCF \geq 0$.
- Modelling can be performed on EAD directly or through CCF.

Our Approach

- Modelling CCF directly is challenging due to instability caused by small/zero headroom, i.e. when $L_0 - B_0 < \epsilon$.
- EAD typically exhibits very long tails. Modelling Usage at Default, $UAD = EAD/L_0$, means estimates are default weighted and not exposure weighted. Referred to as the *momentum* approach.
- We consider a zero-one inflated beta regression model for CCF.
- We shift and scale the beta-distributed random variable to obtain a parametric distribution for UAD.
- This enables us to model UAD directly but also constrains the implicit modelled values for CCF to the unit interval, ensuring that $UAD \geq B_0/L_0 :=$ Usage Before Default (UBD).
- We introduce a margin of conservatism by centering the prior distribution for mean CCF on the standardised value of 75%.
- Models are implemented in the probabilistic programming language `stan` which provides efficient MCMC sampling from the joint posterior distribution of the model parameters.

Data Sources

- We source Corp & FI data from GCD.
- We clean the data keeping only facilities where:
  - there is positive headroom,
  - balance at default is greater than one year prior,
  - balance at default does not exceed limit one year prior.
- This leaves 521 facilities for modelling (summarised in Table →).Note Long tails.
- We source GDP data for different regions (description below ↓).

Table: Observed summary statistics for Balance 1ypd, CCF and EAD

	Balance 1ypd	CCF	EAD
Min.	0	0.000	0
1st Qu.	0	0.032	2031
Median	49	0.516	188518
Mean	3336193	0.501	5094438
3rd Qu.	702088	0.997	2755172
Max.	125296671	1.000	143903572

Table: Description of macro-economic variables used for modelling

Region	Source	Description
UK	ONS	U.K.: Gross Domestic Product, Current Prices (SA, Y/Y % chg)
EA19	EUROSTAT	EA19: Real Gross Domestic Product (SWDA, Y/Y %)
US	BEA	Gross Domestic Product: Percent Change (SAAR, %)

We calibrate the model to UAD

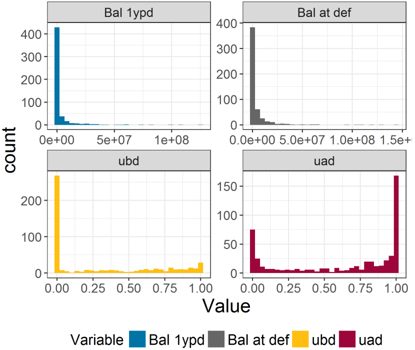

Figure: Histograms of Balance at one year prior to default, balance at default, UBD and UAD

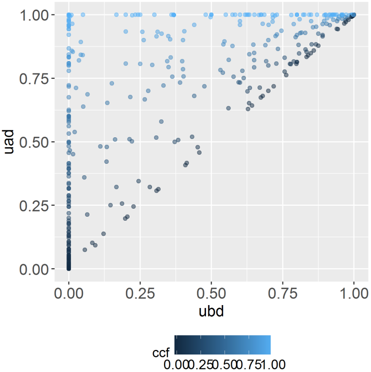

Figure: UAD against UBD. Lighter coloured circles indicate higher CCF.

BARCLAYS

Potential Model Drivers

- We define classes for CCF: (0, 0.5], (0.5, 1), one and zero.
- The table shows average UBD, UAD and GDP by CCF class.
- Data indicates that UBD is on average higher for larger values of CCF.
- Mean GDP is on average higher where CCF is one compared to when CCF is zero.

Table: Mean UBD, UAD and GBP by CCF class: continuous in (0, 0.5], (0.5, 1), one or zero

CCF class	Count	mean UBD	mean UAD	mean GDP
(0,0.5]	164	0.216	0.344	1.940
(0.5,1)	143	0.373	0.867	2.302
one	120	0.340	1.000	1.296
zero	94	0.306	0.306	3.138

Model Parameterisation - Mixture of Beta and Binomial

- We follow Ospina and Ferrari (2012) and induce zero-one inflation via a mixture of Bernoulli and beta distributions.
- Letting $y = CCF$, the density for y is

$$f_{CCF}(y; \mu, \phi, \pi, \theta) = \pi Ber(\theta) + (1-\pi) Beta(\mu, \phi)$$

- Parameters π and θ represent

$$\begin{aligned} \pi &= P(y \in \{0,1\}) \\ \theta &= P(y = 1 | y \in \{0,1\}) \end{aligned}$$

- These are related to $p_0 = P(y = 0)$ and $p_1 = P(y = 1)$ as follows:

$$\begin{aligned} p_0 &= P(y = 0|y \in \{0,1\})P(y \in \{0,1\}) + P(y = 0|y \in (0,1))P(y \in (0,1)) \\ &= (1-\theta)\pi \\ p_1 &= P(y = 1|y \in \{0,1\})P(y \in \{0,1\}) + P(y = 1|y \in (0,1))P(y \in (0,1)) \\ &= \theta\pi. \end{aligned}$$

- We recover π if we compute $p_0 + p_1$.

Model Parameterisation - Shifted and scaled beta distribution

- Suppose X has a beta(μ, ϕ) distribution with $E(X) = \mu \in (0, 1)$ and $Var(X) = \mu(1 - \mu)/(\phi + 1) > 0$.

- Let $Y = a + (b - a)X$. The support of Y is (a, b) and the density of Y is

$$f_Y(y; \mu, \phi, a, b) = \begin{cases} \frac{1}{B(\mu\phi, (1-\mu)\phi - 1)} \frac{(y-a)^{\mu\phi - 1}(b-y)^{(1-\mu)\phi - 1}}{(b-a)^{\phi - 1}}, & y \in (a, a+b) \\ 0 & \text{otherwise} \end{cases}$$

- We use notation $Y \sim Beta_{(a,b)}(\mu, \phi)$ to indicate the support of Y is (a, b).

- If $CCF \sim Beta_{(0,1)}(\mu, \phi)$, then

$$UAD = UBD + CCF(1 - UBD) \sim Beta_{(ubd,1)}(\mu, \phi).$$

- This means we can compute the density, and hence the likelihood, of a model for UAD.

BARCLAYS

Marginal Expectations

- The parameter μ represents the conditional expectation of CCF given CCF is in the open unit interval.

$$\mu = E\{CCF|CCF \in (0,1)\}$$

- The marginal expectation of CCF is

$$\begin{aligned} E(CCF) &= E(CCF|CCF=0)P(CCF=0) + \\ &\quad E(CCF|CCF \in (0,1))P(CCF \in (0,1)) + \\ &\quad E(CCF|CCF=1)P(CCF=1) \\ &= \mu(1-\pi) + \theta\pi \end{aligned}$$

- Conditional mean of UAD is

$$E\{UAD|UAD \in (UBD,1)\} = UBD + (1-UBD)\mu$$

- The marginal expectation of UAD is

$$E(UAD) = UBD + (1-UBD)\{\mu(1-\pi) + \theta\pi\}$$

Likelihood and Prior Distributions

- We regress each model parameter on a set of covariates via a suitable link function and give vague prior distributions to the regression coefficients:

$$\begin{aligned}
\mu &= \text{inv_logit}(x_i^{\mu}\beta^{\mu}), \quad \beta^{\mu} \sim N(0, 10^4 I) \\
\phi &= \exp(x_i^{\phi}\beta^{\phi}), \quad \beta^{\phi} \sim N(0, 10^4 I) \\
\pi &= \text{inv_logit}(x_i^{\pi}\beta^{\pi}), \quad \beta^{\pi} \sim N(0, 10^4 I) \\
\theta &= \text{inv_logit}(x_i^{\theta}\beta^{\theta}), \quad \beta^{\theta} \sim N(0, 10^4 I)
\end{aligned}$$

- The complete set of model parameters is $\Omega = (\mu, \phi, \pi, \theta, \beta)$ where $\beta = (\beta^{\mu}, \beta^{\phi}, \beta^{\pi}, \beta^{\theta})$.

- The likelihood function can be written as

$$\begin{aligned}
L(\Omega; UAD, UBD) &= \prod_{i=1}^{n} \{(1-\theta_i)\pi_i\}^{\mathbb{I}(UAD_i = UBD_i)} (\theta_i \pi_i)^{\mathbb{I}(UAD_i = 1)} \times \\
&\quad \{(1-\pi) Beta_{(UBD_i, 1)}(\mu_i, \phi_i)\}^{\mathbb{I}(UBD_i < UAD_i < 1)}.
\end{aligned}$$

- Strictly, the likelihood is a function of μ, ϕ, π, θ only through β via the deterministic link functions. However, on occasion we will model one or more of μ, ϕ, π, θ directly.

BARCLAYS

Posterior Distributions

- The posterior distribution for Ω is given by

$$f(\Omega|UAD, UBD) \quad \propto \quad L(\Omega; UAD, UBD)f(\Omega)$$

- Given N post burn-in samples from $f(\Omega|UAD, UBD)$ the estimates for CCF_i and UAD_i are calculated as follows:

$$\widehat{CCF_i} \quad = \quad \frac{1}{N}\sum_{t=1}^{N} \mu_i^{(t)}(1-\pi_i^{(t)}) + \theta_i^{(t)}\pi_i^{(t)}$$

$$\widehat{UAD_i} \quad = \quad \frac{1}{N}\sum_{t=1}^{N} UBD_i + (1-UBD_i)\{\mu_i^{(t)}(1-\pi_i^{(t)}) + \theta_i^{(t)}\pi_i^{(t)}\}$$

- We can obtain 95% credible interval from the samples $CCF^{(t)}$ and $UAD^{(t)}$.
- For readability, change notation to let $y = UAD$ and $x = UBD$.
- Prediction for out-of-sample data $y_0 = UAD_0$, requires sampling from the posterior predictive distribution.

$$f(y_0|x_0, y, x) = \int f(y_0|x_0, \Omega)f(\Omega|y, x)d\Omega$$

- If $\Omega^{(t)} \sim f(\Omega|y, x)$ then $y_0^{(t)} \sim f(y_0|x_0, \Omega^{(t)})$ is a draw from the posterior predictive distribution $f(y_0|x_0, y, x)$.

Candidate models

- We look at different regression models for the parameters using available covariates UBD and GDP.
- Where we do not have a regression model for a parameter, we estimate it directly rather than through β_0. For example, we do not place a regression model on π, instead the prior is $\pi \sim U(0, 1)$.
- We use the notation $\mu(ubd, gdp)$ to indicate that we model $\mu = \text{inv\_logit}(\beta_0 + \beta_1 ubd + \beta_2 gdp)$.
- A benchmark model for UAD is the Tobit model:
  - The Tobit model assumes a normal distribution but only observed for the unit interval. Point masses at zero and one.
  - We model the mean and variance of the underlying normal distribution. There is no closed form solution for the MLEs. Parameter estimates are obtained by Newton-Raphson optimisation.

Model Selection Results

- Models are fit to 80% of the data, holding out 20% for testing.
- Inference is based on 3000 post burn-in samples generated from three parallel MCMC chains.
- Model performance is judged by WAIC (widely-applicable IC), which is appropriate for Bayesian analysis (Watanabe, 2010), and prediction RMSE.
- The best performing model is M7 with $\mu(ubd), \theta(gdp)$.
- The best Tobit model has $\mu(ubd, gdp), \sigma^2(ubd, gdp)$ where σ^2 is modelled via a log-link.

Table: WAIC and RMSE for a selection of models

Model	Description	WAIC	RMSE
M1	$\mu(ubd)$	892.1	0.341
M2	$\mu(gdp)$	906.6	0.348
M3	$\mu(ubd, gdp)$	893.7	0.339
M4	$\theta(ubd)$	906.9	0.344
M5	$\theta(gdp)$	895.7	0.340
M6	$\mu(ubd), \theta(ubd)$	893.6	0.339
M7	$\mu(ubd), \theta(gdp)$	882.8	0.336
M8	$\mu(ubd), \phi(ubd)$	894.1	0.339
M9	$\mu(ubd), \phi(gdp)$	893.6	0.339

The RMSE for the best Tobit model is 0.346.

Parameter Estimates for Best Model

Table: Parameter estimates for Model M7: $\mu(ubd), \theta(gdp)$

parameter	mean	sd	2.5%	97.5%	$\hat{R}$
$\mu(\beta_0)$	-0.066	0.079	-0.221	0.096	0.999
$\mu(\beta_1)$	0.326	0.092	0.145	0.513	0.999
ϕ	1.066	0.048	0.977	1.170	0.999
π	0.398	0.025	0.353	0.448	0.999
$\theta(\beta_0)$	0.426	0.170	0.101	0.756	1.002
$\theta(\beta_1)$	-0.578	0.166	-0.921	-0.281	0.999

- $\mu(\beta_1)$ indicates a significant correlation between the logit of conditional expectation of CCF and UBD.
- $\theta(\beta_1)$ indicates that GDP is a driver for θ and hence p_0 and p_1.
- $\hat{R}$ is an estimate of the potential scale reduction factor which is a well-known convergence statistic. Values below 1.1 are used to indicate convergence of the Markov chain to the posterior distribution of the parameter. **(?)**

Box plots of facility-level effects

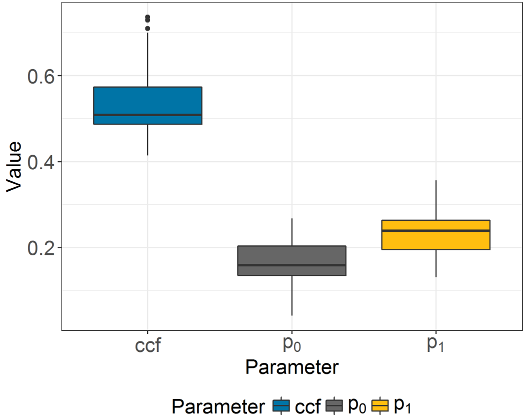

Figure: Boxplots of parameter estimates for CCF, p_0 and p_1 for model M7

- p_0 and p_1 vary over i due to dependence on θ.

Conservative Prior Distribution

- We embed conservatism by adjusting the prior distribution for the conditional mean of CCF, μ.
- For the preferred model we have $\mu = \text{inv\_logit}(\beta_0 + \beta_1 UBD)$, where $\beta_k \sim N(m_k, v_k)$ for $k = 0, 1$.
- We set m_0 so that the prior mean is equal to the standardised value of 75% $m_0 = \text{logit}(0.75) = \log(0.75/(1 - 0.75)) = \log(3)$.
- We set v_0 so that the β_0 is significantly different from 0, $P(\beta_0 < 0) = \alpha/2$, where α is 1 - confidence level, the strength of belief in the standardised value.
- Re-arranging gives prior standard deviation $\sqrt{v_0} = m_0/\Phi^{-1}(\alpha/2)$.
- All modelled CCF values receive an uplift. With 95% confidence ($\alpha = 5\%$) the average uplift is 1%, 99% confidence gives 1.5%.
- Strong prior belief does not strongly effect inference as the data overwhelms the prior.

Table: Parameter estimates Regulatory model

parameter	mean	sd	2.5%	97.5%	$\hat{R}$
$\mu(\beta_0)$	-0.042	0.081	-0.201	0.107	0.999
$\mu(\beta_1)$	0.332	0.094	0.159	0.514	0.999
ϕ	1.062	0.049	0.968	1.160	0.999
π	0.400	0.025	0.352	0.450	1.001
$\theta(\beta_0)$	0.486	0.158	0.177	0.796	0.999
$\theta(\beta_1)$	-0.573	0.172	-0.908	-0.253	0.999

Summary

- We have described a fully Bayesian approach to modelling EAD.
- This properly accounts for uncertainty in parameter estimation when predicting new values.
- By shifting and scaling a beta distribution we build a parametric regression model for UAD.
- We show how the prior distribution can be used to embed conservatism into model predictions.

Future Work

- A simulation study demonstrating parameter recovery.
- Broader data collection for covariates.
- Full k-fold cross-validation.
- Wider model comparison.
- How to handle balances in excess of limits.

References

Raydonal Ospina and Silvia L.P. Ferrari. A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis*, 56(6):1609 – 1623, 2012. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2011.10.005. URL `http://www.sciencedirect.com/science/article/pii/S0167947311003628`.

Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual, Version 2.18.0*, 2018. URL `http://mc-stan.org`.

Edward NC Tong, Christophe Mues, Iain Brown, and Lyn C Thomas. Exposure at default models with and without the credit conversion factor. *European Journal of Operational Research*, 252(3):910–920, 2016.

Sumio Watanabe. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(Dec):3571–3594, 2010.

Fang Liu with contributions from Yunchuan Kong. *zoib: Bayesian Inference for Beta Regression and Zero-or-One Inflated Beta Regression*, 2018. URL `https://CRAN.R-project.org/package=zoib`. R package version 1.5.1.

stan

- `stan` is a probabilistic programming language providing efficient MCMC sampling from posterior distributions.
- There are 6 blocks: `functions`, `data`, `parameters`, `model`, `transformed parameters` and `generated quantities`.

```
functions {
  real generalised_beta_log_pdf(real y, real p, real q, real a) {
  return beta_lpdf(y | p, q) + (p - 1)*(log(y - a) - log(y)) - pow(1 - a, p + q - 1);
...
  }
data {
  int n;
  vector[n] x; //ubd
  vector<lower=0, upper=1>[n] y; //uad
...
}
parameters {
  real<lower=0, upper=1> Pi;
  real<lower=0> phi;
  vector[2] coef_theta;
  vector[2] coef_mu;
  }
```

stan

```
transformed parameters {
  vector<lower=0>[n] p;        // scale 1 of beta(p,q)
  p = mu * phi;
...
}
model {
  Pi ~ uniform(0, 1); coef_theta ~ normal(0, sqrt(10000));
  coef_mu ~ normal(0, sqrt(10000)); phi ~ lognormal(0, sqrt(10000));
  for (i in 1:n) { // zero-one inflated beta likelihood
    if (y[i] == x[i]) {
      target += log(Pi) + log1m(theta[i]);
    } else if (y[i] == 1) {
      target += log(Pi) + log(theta[i]);
    } else {
      target += log1m(Pi) + generalised_beta_log_pdf(y[i] , p[i], q[i], x[i]);
    }
  }
}
generated quantities {
  vector[n] ccf_marg_exp;
  for (i in 1:n) ccf_marg_exp[i] = Pi*theta[i] + (1-Pi)*mu[i];
...
  }
```