Estimating the asset correlation in the single risk factor model on Dutch mortgage data

P. van den Berg, A. Colangelo

6 August, 2014

Introduction

Banks calculate the required capital for unexpected credit losses using a prescribed formula,

$$\mathrm{RWA} \propto (\mathrm{PD}_{\mathrm{stressed}} - \mathrm{PD}_{\mathrm{lta}})$$

where $\mathrm{PD}_{\mathrm{lta}}$ is the bank's estimate of the long term average yearly default probability of an exposure, typically conditioned on counterparty and loan-specific factors and PD_{stressed} is the 'stressed' PD, which corresponds to the 99.9^{th} percentile of the distribution of an assumed single risk factor representing systemic risk, i.e., a risk factor which determines the amount of correlation between defaults and which is assumed to fluctuate year by year.

Model specification

Our data consists of $N = \sum_{b,t} N_{bt}$ observations of binary variable

$$D_i^{(bt)}, i = 1 \ldots N_{bt}, b = 1 \ldots B, t = 1 \ldots T$$

The observations are segmented over B buckets which represent some ordering over the average default rate, i.e., a PD model. This segmentation we regard as given. The single risk factor model is specified as follows: the likelihood of finding k_{bt} defaults out of N_{bt} observations given the value of the systemic factors y_t, the long-term average default rates λ_b and the asset correlation ρ, is given by:

$$P(k_{bt}|\rho, \lambda_b, y_t, N_{bt}) = \binom{k_t}{N_t} G_{bt}^{k_{bt}} (1 - G_{bt})^{N_{bt} - k_{bt}}$$

where the G_{bt} are defined as

$$G_{bt}(y_t; \lambda_b, \rho) = \Phi(\frac{\Phi^{-1}(\lambda_b) - \sqrt{\rho} y_t}{\sqrt{1-\rho}})$$

The systemic factors $\mathbf{y}$ are assumed to be independently normally distributed,

$$\mathbf{y} \sim \varphi(y_1) \dots \varphi(y_T)$$

For the sake of simplicity, we choose independent priors for the λ_b and ρ. We take the Beta distribution as the marginal prior for λ, since for $\rho = 0$ this is a natural choice of prior for the λ_b.

$$P(\lambda|\alpha, \beta, I) = \frac{1}{\mathrm{Beta}(\alpha, \beta)} \lambda^{\alpha-1}(1-\lambda)^{\beta-1}$$

For ρ we choose a uniform prior on $[0, 1]$,

$$P(\rho) = \begin{matrix} 1 & 0 <= \rho <= 1 \\ 0 & \text{otherwise} \end{matrix}$$

Using Bayes theorem for inverting conditional probabilities, the joint posterior for λ and ρ is then

$$P(\lambda, \rho, \mathbf{y}|\mathbf{k}, \mathbf{N}, I) \propto P(\mathbf{k}|\rho, \lambda, \mathbf{y}, \mathbf{N}, I) P(\rho) \prod_b P(\lambda_b) \prod_t P(y_t)$$

The marginal posterior $P(\lambda, \rho|k, N, I)$ is found by integrating out the unobserved variable y.

We implement this model in Stan [@stan-software:2014]. See the appendix for the complete model code.

Data

We use a dataset including yearly default incidences for Dutch mortgages from September 2007 to September 2013. The data was collected directly from Dutch banks for the purposes of risk model validation, and covers $> 95\%$ of total bank exposures to Dutch mortgages.

As a first step, we define a constant categorization (‘bucketing’) of the data such that the average yearly default rate over the available history differs significantly per bucket (category). [...]

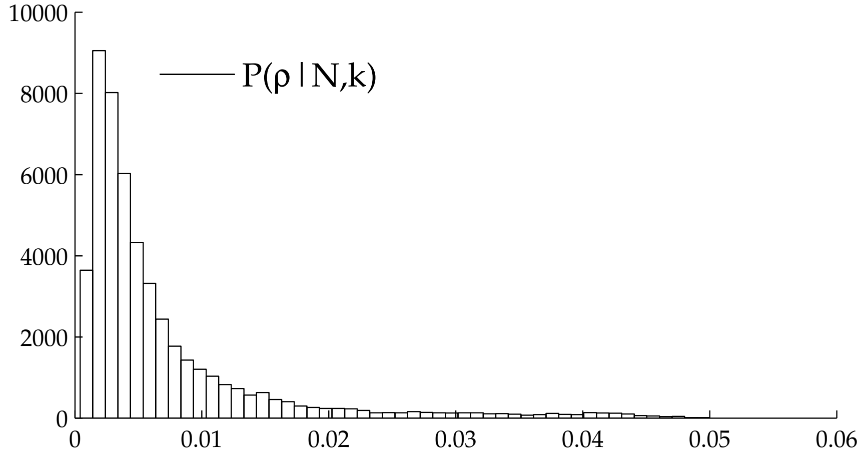

Figure 1: Posterior of ρ

Results

See figure.

Appendix: Stan implementation

```
data {
  int<lower=0> T; // time periods
  int<lower=0> B; // uniform risk categories (PD buckets)
  int k[B,T];
  int N[B,T];
}
parameters {
  real nu[B]; // nu = normal_cdf_inv(lambda)
  real<lower=0,upper=1> rho;
  real y[T];
}
transformed parameters {
  real<lower=0,upper=1> G[B,T];

  for (t in 1:T)
    for (b in 1:B)
      G[b,t] <- normal_cdf( (nu[b] - sqrt(rho) * y[t]) / sqrt(1-rho) , 0.0, 1.0);

}
```

```
model {
  for (b in 1:B)
   increment_log_prob(beta_log(normal_cdf(nu[b], 0.0, 1.0), 1.0, 1.0));
  rho ~ uniform(0.0,1.0);
  y ~ normal(0.0,1.0);

  for (t in 1:T)
  for (b in 1:B)
  k[b,t] ~ binomial(N[b,t], G[b,t] );
}

generated quantities {
  real lambda[B];
  for (b in 1:B)
    lambda[b] <- normal_cdf(nu[b], 0.0, 1.0);
}
```