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,

where 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 is the ‘stressed’ PD, which corresponds to the 99.9 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 observations of binary variable

The observations are segmented over 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 defaults out of observations given the value of the systemic factors , the long-term average default rates and the asset correlation , is given by:

$$
P(k\_{bt}|\rho,\lambda\_b,y\_t,N\_{bt})=\left(\begin{matrix} k\_t \\ N\_t\end{matrix}\right)
G\_{bt}^{k\_{bt}} (1-G\_{bt})^{N\_{bt}-k\_{bt}}
$$

where the are defined as

The systemic factors are assumed to be independently normally distributed,

For the sake of simplicity, we choose independent priors for the and . We take the Beta distribution as the marginal prior for , since for this is a natural choice of prior for the .

For we choose a uniform prior on ,

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

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

The marginal posterior is found by integrating out the unobserved variable .

We implement this model in Stan [@stan-software:2014]. See the [appendix](#app) 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 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). […]

## Results

![](data:application/pdf;base64,)

Posterior of

See [figure](#fig:rhofig).

# 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);  
}