Leveraging Probability of Default Models for Bayesian Inference of Default Correlations

Miguel Biron[1], Víctor Medina[1]

November 2017

[1]Superintendency of Banks and Financial Institutions (SBIF), Santiago, Chile

Summary

Motivation: capital requirements for banks

The Single Risk Factor (SRF) model

Previous work

Two-stage inference procedure

Simulations

Experiments

Conclusions

Summary

We present a method to infer the default correlation between loans in a Single Risk Factor (SRF) model for a loan portfolio

We propose a two-stage procedure that takes advantage of Probability of Default (PD) models already trained for the portfolio

We study the performance of this approach both with simulated data and on a dataset of a portfolio of mortgages

Capital requirements under Basel II

Banks using the Internal Ratings Based (IRB) approach, compute the risk weights of their exposures using (BCBS 2006):

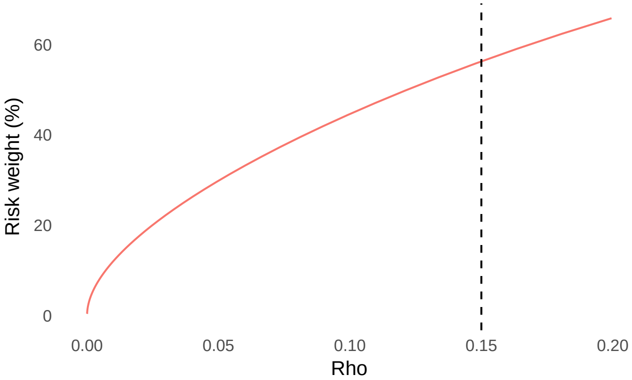
$$\text{Risk Weight} = 12.5 \cdot LGD \left[\Phi\left(\frac{\Phi^{-1}(PD) + \sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}} \right) - PD \right]$$

where

- α: confidence level (must be 0.999 for all exposures)
- PD: probability of default (computed by the bank)
- LGD: loss given default (could be computed by the bank)
- ρ: default correlation parameter (imposed by the regulator)
    - For residential mortgages exposure, $\rho = 0.15$

Risk weights are very sensitive to changes in ρ

Assuming $PD = 0.20$ and $LGD = 0.10$:



Empirical assesments of ρ

In spite of the importance of ρ, empirical assessments of its magnitude are rare

In this work, we propose a method for inferring this parameter from data on historical defaults

These data correspond to binary observations y_{it}, such that

$$y_{it} = \begin{cases} 1 & \text{client } i \text{ defaults at time } t \\ 0 & \text{otherwise} \end{cases}$$

However, in order to describe this method, we first need a probabilistic definition of the Single Risk Factor (SRF) model for credit portfolios

Merton's model

Merton (1974) described a method for pricing corporate debt instruments

In this setup, a firm i defaults at time t if its liabilities D_i are larger than the value of its assets A_{it}

Merton postulated a stochastic process for the value of these assets, which after some manipulation can be expressed as

$$\log(A_{it}) = \log(A_{i0}) + \left(\mu_i - \frac{1}{2}\sigma_i^2\right) t + \sigma_i \sqrt{t} x_{it} \tag{1}$$

where $x_{it} \sim N(0, 1)$.

Merton's model

Thus, the firm defaults iif

$$y_{it} = 1 \Leftrightarrow x_{it} < \frac{\log \frac{D_i}{A_{it}} - (\mu_i - \frac{1}{2}\sigma_i^2)t}{\sigma_i \sqrt{t}} \triangleq x_{it}^* \tag{2}$$

where x_{it}^* is known as the Distance to Default. Moreover, the probability of default can be expressed as

$$PD_{it} \triangleq \mathbb{P}(y_{it} = 1) = \mathbb{P}(x_{it} \leq x_{it}^*) = \Phi(x_{it}^*) \tag{3}$$

where $\Phi(\cdot)$ is the standard normal CDF.

Extending Merton's model to retail exposures

If one knows the PD, but does not know any other firm characteristic, then one can recover the distance to default by inverting the above formula

$$x^*_{it} = \Phi^{-1}(PD_{it}) \tag{4}$$

Although trivial, this last step is crucial because it allows us to extend the model to counterparties other than corporations, where assets are not the main driver of default (Rösch and Scheule 2004)

Vacicek's model

Vasicek (1987) introduced dependency to Merton's model. This is achieved by the following decomposition of x_{it}

$$x_{it} = \sqrt{\rho} M_t + \sqrt{1-\rho}\varepsilon_{it} \tag{5}$$

where $M_t \sim N(0,1)$, $\varepsilon_{it} \overset{iid}{\sim} N(0,1)$ and $M_t \perp \varepsilon_{it} \, \forall i, t$

It can be shown that this decomposition preserves $x_{it} \sim N(0,1)$, and, more importantly, that

$$Cov(x_{it}, x_{jt}) = \rho, \, \forall i \neq j, t \tag{6}$$

Probabilistic model of default data

From the previous equations, we can "marginalize out" the x_{it}'s by noting that

$$\mathbb{P}(y_{it} = 1 | M_t, \rho) = \mathbb{P}(x_{it} \leq x_{it}^* | M_t, \rho) = \Phi\left(\frac{x_{it}^* - \sqrt{\rho} M_t}{\sqrt{1-\rho}}\right) \tag{7}$$

Using this fact, we arrive at a hierarchical probabilistic model of the SRF portfolio

$$\begin{aligned} \rho &\sim U(0,1) \\ M_t &\overset{iid}{\sim} N(0,1) \\ y_{it} | M_t, \rho &\overset{indep}{\sim} \text{Bernoulli}\left(\Phi\left(\frac{x_{it}^* - \sqrt{\rho} M_t}{\sqrt{1-\rho}}\right)\right) \end{aligned} \tag{8}$$

Bayesian inference for the SRF model

If we were given the true distances to default x^*_{it}, then we could infer $\boldsymbol{\theta} = (M_t, \rho)$ by using Bayes' Theorem:

$$\underbrace{p(\boldsymbol{\theta}|\mathbf{Y})}_{\text{Posterior}} = \frac{p(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{Y})} \overset{\boldsymbol{\theta}}{\propto} \underbrace{p(\mathbf{Y}|\boldsymbol{\theta})}_{\text{Likelihood}} \underbrace{p(\boldsymbol{\theta})}_{\text{Prior}} \tag{9}$$

where $\mathbf{Y} = \{y_{it}\}$ is a matrix of default observations of size $T \times N$

Alas, the true distances to default are unknown...

Previous work

- Rösch and Scheule (2004)
  - Proposes a Maximum Likelihood (ML) inference procedure for ρ
  - Estimated jointly with a generalized linear model (GLM) for the PD of the portfolio
  - Since regression is at the portfolio level, it does not include individual-level covariates (only macroeconomic)
  - Applications to portfolios of mortgages ($\rho \approx 0.0028$), credit cards ($\rho \approx 0.0066$) and other consumer loans ($\rho \approx 0.0044$)
- McNeil and Wendin (2007)
  - Suggests a Bayesian GLM mixture model (GLMM) approach
  - ρ becomes a function of the variance of the latent factor
  - Application to a portfolio of corporate bonds yields $\rho \approx 0.075$
  - Includes only a few covariates

Previous work

- J. Crook and Bellotti (2012)
  - Maximum Likelihood inference of a GLMM model
  - Applied to credit card portfolio
  - Includes lots of covariates (account, client and macroeconomic level)
  - Interesting finding: ρ is <u>lower</u> in times of economic distress ($\rho \approx 0.0024$ versus $\rho \approx 0.0003$)

It is important to note that all of these methodologies require estimating a PD model in order to infer ρ

Summary

Motivation: capital requirements for banks

The Single Risk Factor (SRF) model

Previous work

Two-stage inference procedure

Simulations

Experiments

Conclusions

Incorporating the PD model

In order to deal with the unknown x_{it}^*, we propose a two-stage procedure that can be stated as

$$\hat{x}_{it}^* = \Phi^{-1}(\widehat{PD}_{it}) \tag{10}$$

where $\hat{x}_{it}^*$ is a point estimate of the true x_{it}^*, and $\widehat{PD}_{it}$ the output of the PD model for the (i, t) observation

In other words, our approach is to use a "plug-in" estimator of the distances to default, given by a PD model previously fitted for the portfolio

Comments on the two-stage procedure

Two-stage methods are common in the estimation of copulas (Xu 1996; Joe 2005). The two stages are:

1. Infer marginal distribution of components assuming independence (in our case, the PD model)
2. Infer copula parameters, given the estimates for the marginal distributions (in our case, ρ)

Benefits of the two-stage procedure

- We can take advantage of PD models because they are common components of risk management frameworks (Hand and Henley 1997; J. N. Crook, Edelman, and Thomas 2007)
  - No need to re-estimate a PD model just to infer ρ
- We only need to know the point estimates for the PD and not the whole model
  - Hence, we can accommodate any PD model
  - This is important, since non-linear and non-parametric methods with many covariates have shown better performance over simpler linear alternatives (Lessmann et al. 2015)

Drawback of the two-stage procedure

Inferences on ρ could be sensitive to small errors in the point estimates of the PD. To correct this, we propose a "robust" alternative:

$$\begin{aligned}
\rho &\sim U(0,1) \\
\beta &\sim U(-\infty,\infty) \\
M_t &\overset{iid}{\sim} N(0,1) \\
y_{it}|M_t,\rho,\beta &\overset{indep}{\sim} \text{Bernoulli}\left(\Phi\left(\frac{\beta\hat{x}^*_{it}-\sqrt{\rho}M_t}{\sqrt{1-\rho}}\right)\right)
\end{aligned} \tag{11}$$

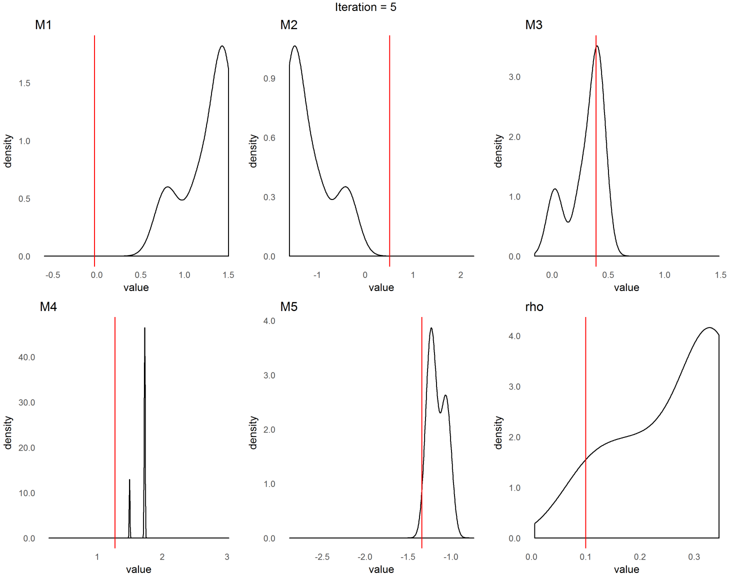
Here, instead of passing the point estimates directly, we fit a GLM model with the point estimate as the sole predictor

Note that when $\beta=1$, we recover the "naive" approach (nested)

Computational aspects

We implement the models in Stan, a probabilistic programming language that uses a Hamiltonian Monte Carlo algorithm to draw samples from the posterior distribution (Carpenter et al. 2016)

Hamiltonian Monte Carlo algorithm



Model coherence check

If we generate data according to the SRF model, we should be able to recover the parameters that generated those data with our inference procedures

In particular, Credible Intervals (CI) of given nominal coverage α should achieve a true coverage of α in repeated samples

Following Si et al. (2015), we:

- Sample $K = 500$ trajectories for a toy portfolio with $T = 20$ and $N = 200$ using the SRF model
- For each portfolio, fit the models and build CI's with $\alpha \in \{50\%, 80\%, 95\%\}$
- Finally, assess the empirical coverages of the CI's

Robustness to distortions in the PDs

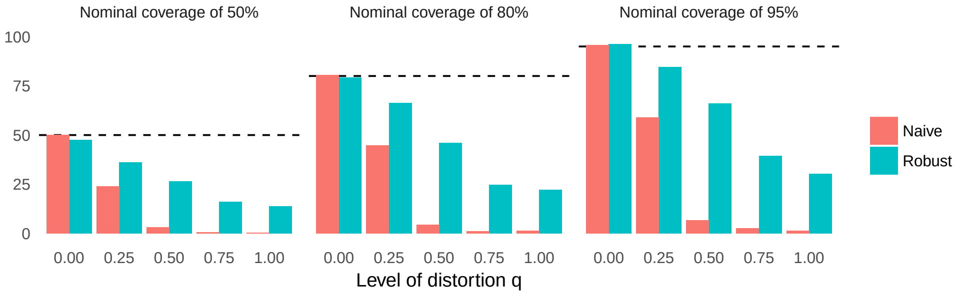
The SRF model assumes the PDs to be exogenous. Thus, we generate these quantities once from a Beta distribution and feed them to the models for all the trajectories

However, to test the resilience of our methods, we also try to infer ρ using distorted versions of the PDs

We control this distortion with a parameter $q \in [0, 1]$, such that

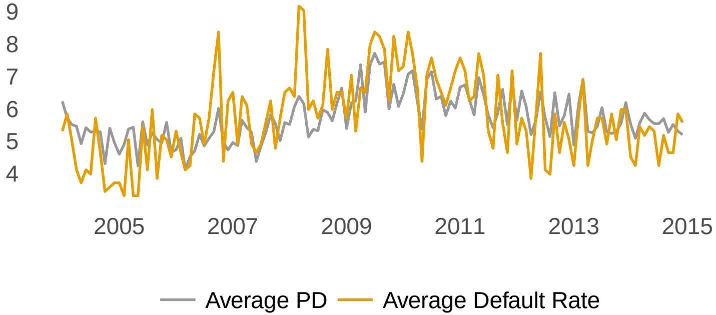
- $q = 0$: no distortion. We infer ρ using the true PDs (best case scenario)
- $q = 1$: full distortion. We infer ρ using noise as the PDs (worst case scenario)

True versus nominal coverage rates



- When $q = 0$, both approaches achieve the true coverage
- On the other hand, as q grows, both methods begin to fail
  - However, the robust method fails considerably less

Application to a mortgage portfolio

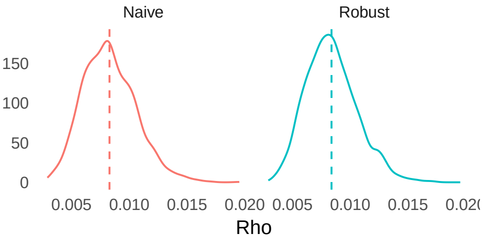


Application to a mortgage portfolio from Chilean banks, for which a PD model exists at SBIF (Biron and Urbina 2017)

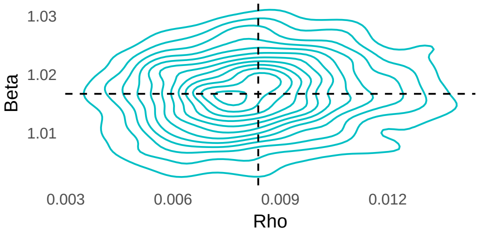
Sample size is $T = 132$ months and $N = 754$ individuals per month

Posterior distribution of ρ and β

Kernel density estimators from MCMC posterior distribution samples



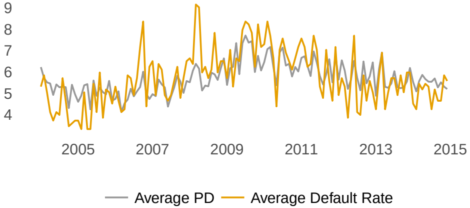
(a) $p(\rho|\mathbf{Y})$



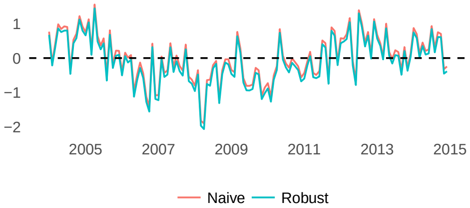
(b) $p(\rho, \beta|\mathbf{Y})$ (robust only)

- Output of both methods is similar
- Nevertheless, the region $\beta > 1$ shows most of the mass, which implies a small but significant correction by the robust method

Posterior distribution of the latent factor



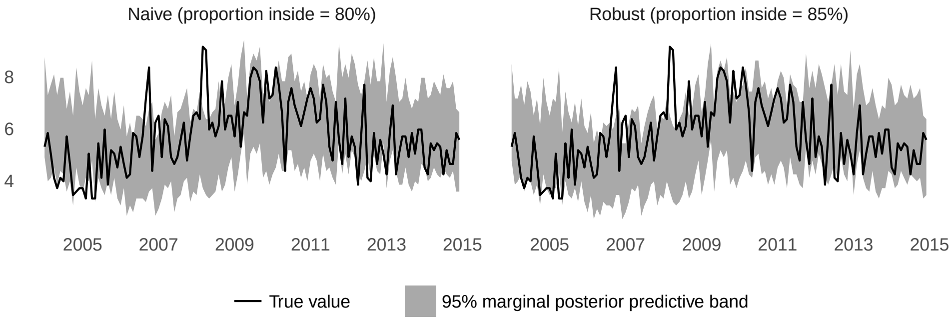
(a) Monthly default rate and average PD



(b) Posterior mean of M_t

- M_t tends to capture the volatility in the default rate that is not explained by the PD model
- Output of both methods is similar

Posterior predictive checks



- Using simulations from the posterior distribution, we re-generate the time series for the average default rate
- The robust approach achieves a better coverage, although none of them achieves the target of 95%

Comparison using an information criterion

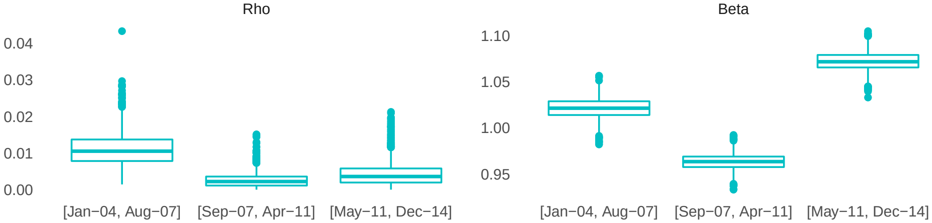
- The Watanabe-Akaike Information Criterion (WAIC) evaluates the ability of models to generalize to new data (bias-variance trade-off)

Model	Bias^{-1} (1)	Variance (2)	WAIC = (1) - (2)
Naive	-13770	63	-13833
Robust	-13764	65	-13829
Fit PD	-13697	151	-13849

- The robust approach achieves the highest WAIC
- Fitting a new PD model jointly with the correlation shows the worst WAIC
  - Its low bias does not compensate for the additional variance introduced by the parameters of the PD model

Constant ρ?

We investigate the assumption of constant ρ by applying the robust method to 3 disjoint windows of time of equal length



- The posterior distribution of both parameters (ρ, β) seems to be non-stationary
- Interestingly, we recover the finding by J. Crook and Bellotti (2012): ρ is lower when the default rate is higher

Conclusions and further work

- Robust outperforms naive in simulations and real datasets
- Both methods outperform re-estimating the whole PD model
- Assumption of constant ρ seems unlikely $\rightarrow$ lends support to improving the SRF model
  - For example, ρ having different probable "states" and model the probability of each state dynamically (Hidden Markov Model)
  - Other approach is to think of clusters with different sensitivity to shocks (ie, different ρs), which might increase or decrease in proportion dynamically (mixture of distributions)

Thank you

References I

BCBS. 2006. *International Convergence of Capital Measurement and Capital Standards: A Revised Framework*. Bank for International Settlements.

Biron, Miguel, and Miguel Urbina. 2017. "Un Indicador Prospectivo de Riesgo de Crédito de Personas Con Enfoque Sistémico." *SBIF Working Paper*.

Carpenter, Bob, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2016. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 20.

Crook, Jonathan N, David B Edelman, and Lyn C Thomas. 2007. "Recent Developments in Consumer Credit Risk Assessment." *European Journal of Operational Research* 183 (3). Elsevier: 1447–65.

Crook, Jonathan, and Tony Bellotti. 2012. "Asset Correlations for Credit Card Defaults." *Applied Financial Economics* 22 (2). Taylor & Francis: 87–95.

Hand, David J, and William E Henley. 1997. "Statistical Classification Methods in Consumer Credit Scoring: A Review." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 160 (3). Wiley Online Library: 523–41.

Joe, Harry. 2005. "Asymptotic Efficiency of the Two-Stage Estimation Method for Copula-Based Models." *Journal of Multivariate Analysis* 94 (2). Elsevier: 401–19.

Lessmann, Stefan, Bart Baesens, Hsin-Vonn Seow, and Lyn C Thomas. 2015. "Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring: An Update of Research." *European Journal of Operational Research* 247 (1). Elsevier:

References II

124–36.

McNeil, Alexander J, and Jonathan P Wendin. 2007. "Bayesian Inference for Generalized Linear Mixed Models of Portfolio Credit Risk." *Journal of Empirical Finance* 14 (2). Elsevier: 131–49.

Merton, Robert C. 1974. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *The Journal of Finance* 29 (2). Wiley Online Library: 449–70.

Rösch, Daniel, and Harald Scheule. 2004. "Forecasting Retail Portfolio Credit Risk." *The Journal of Risk Finance* 5 (2). Emerald Group Publishing Limited: 16–32.

Si, Yajuan, Natesh S Pillai, Andrew Gelman, and others. 2015. "Bayesian Nonparametric Weighted Sampling Inference." *Bayesian Analysis* 10 (3). International Society for Bayesian Analysis: 605–25.

Vasicek, Oldrich. 1987. "Probability of Loss on Loan Portfolio." *Mimeo*. KMV Corporation.

Xu, James Jianmeng. 1996. "Statistical Modelling and Inference for Multivariate and Longitudinal Discrete Response Data." PhD thesis, University of British Columbia.