Estimating default probability and correlation using Stan

Jesús Antonio Piñera Esquivel

Vasicek factor model

Given a credit portfolio with N_t debtors at the start of the period (usually year or month), and D_t defaults at the end of the period, we assume that $D_t \mid N_t, \pi_t \sim Bin(N_t, \pi_t)$. The parameter π_t is called the **conditional default probability**, and it is assumed of the form

$$\pi_t = \Phi\left(\frac{\Phi^{-1}(p) - \sqrt{\rho}Z_t}{\sqrt{1-\rho}}\right)$$

where p and ρ are the (unconditional) default probability and correlation, respectively, and $Z_t \sim N(0,1)$ is called the factor. In this case, the conditional probability π has closed form distribution[1] called the **Vasicek distribution**, denoted as $\pi \sim Vas(p, \rho)$.

[1] Easy to obtain just by solving for Z_t.

Vasicek distribution

The following figure shows the different forms the distribution can take changing both parameters. Note how the parameter p behaves like a mean and ρ as variance parameter[2].

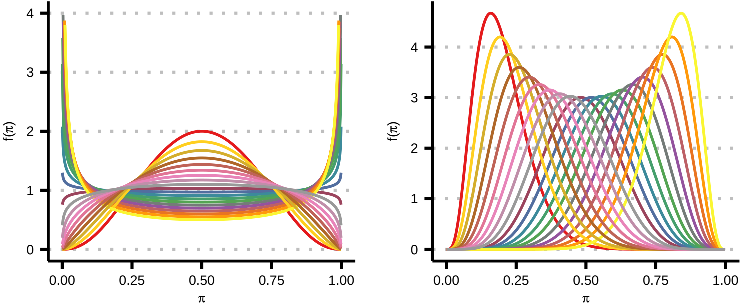

Figure 1: Vasicek density for different values of the default correlation, with fixed default probability in 0.5

[2] Indeed, $E[\pi \mid p, \rho] = p$, while the variance is an increasing function of ρ.

Parameter estimation

Classical methods like MLE actually use the form $D_t \mid Z_t \sim Bin(N_t, \pi(Z_t))$, and calculate the density as

$$f(D_t) = \int_{\mathbb{R}} \binom{N_t}{D_t} \pi(z)^{D_t} (1 - \pi(z))^{N_t - D_t} d\Phi(z)$$

where $\pi(z) = \Phi((\Phi^{-1}(p) - \sqrt{\rho}\, z)/\sqrt{1-\rho})$. Then, one maximizes $\prod_{t=1}^{T} f(D_t; p, \rho)$ to obtain the estimates.

However, the form of the factor model allows for a straightforward Bayesian inference. We have that $D_t \mid \pi_t \sim Bin(N_t, \pi_t)$ and that $\pi_t \mid p, \rho \sim Vas(p, \rho)$. We just have to establish appropriate priors for p and ρ. The full posterior can be written as

$$f(\pi, p, \rho | D, N) \propto \left(\prod_{t=1}^{n} Bin(N_t, \pi_t) \right) \left(\prod_{t=1}^{n} Vas(p, \rho) \right) f(p, \rho).$$

where $D = (D_1, ..., D_n)$ and $N = (N_1, ..., N_n)$.

Prior choosing

First, considering the nature of credit portfolios, we rarely expect p to go beyond 0.5, so most of its mass should be concentrated before that point. Moreover, given that ρ behaves as a variance parameter, we suggest the form $f(p, \rho) = f(p \mid \rho) f(\rho)$. For both densities, we use beta proportion distributions, such that $f(\rho) = BetaP(\mu_\rho, \phi_\rho)$ and $f(p|\rho) = BetaP(\mu_p, a\rho)$ with a a positive number.

We use $\mu_\rho = 0.5$ and $\phi_\rho = 5$ to make the prior weakly informative for ρ. On the other hand, we use $\mu_p = 0.2$ to take into account that default probabilities are usually very low, and $a = 10$ so also the ϕ parameter of the probability is around 5.

Simulation study

Parameters' chains

We simulate the process for 4 different pairs of parameters: $p = 0.01$, $\rho = 0.1$ (LL), $p = 0.01$, $\rho = 0.5$ (LH), $p = 0.05$, $\rho = 0.1$ (HL), $p = 0.05$, $\rho = 0.5$ (HH). The sample size consists of 20 points. The following shows the chains of each parameter, for each series.

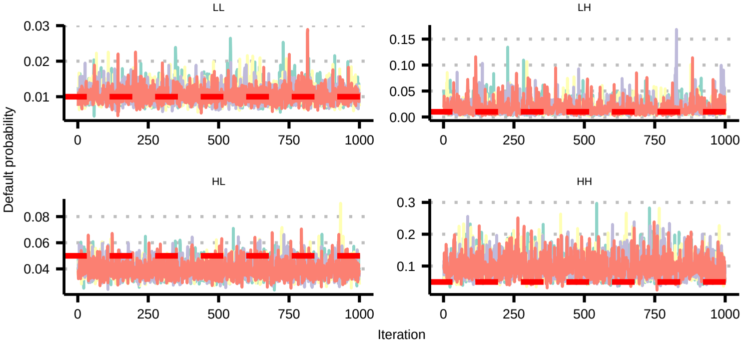

Figure 2: Chains for the default probability for each pair of parameters

Simulation study

Parameters' chains

We can observe that the real parameters (in red) land in the posterior distributions, and for the correlation the mean value is close to the real value.

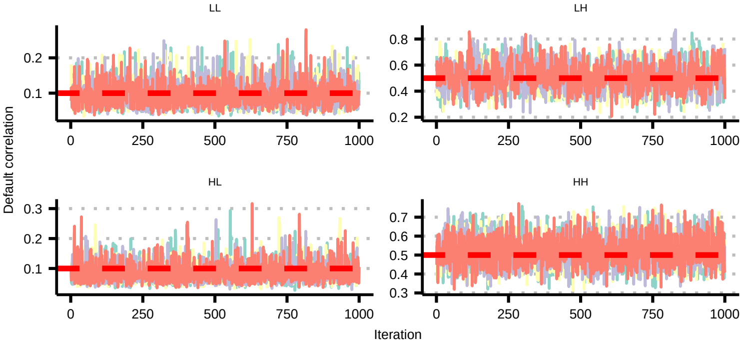

Figure 3: Chains for the default correlation for each pair of parameters

Simulation study

Model fit to data

To assess model appropriateness, we plot the density overlays for each of the series. We can observe that the observed density is contained within the draws' densities.

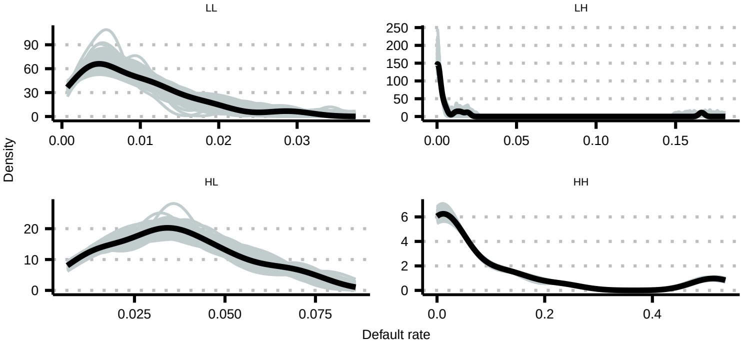

Figure 4: Chains for the default correlation for each pair of parameters

Simulation study

Comparison with classic models[3]

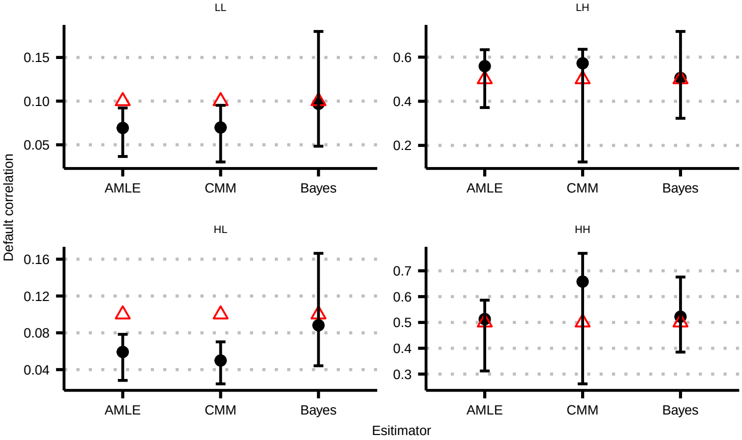

Figure 5: 95% confidence intervals and 95% credibility interval

[3]AMLE is asymptotic MLE, see Gordy (2000); CMM is the AR(1) corrected method of moments, see Frei and Wunsch (2018). 10 thousand bootstrap replicates are used.

Applying the model to S&P corporate default rates

We have 41 months of default rates, for both investment grade (IG) firms and speculative grade (SG) firms. We use this data to fit the factor model and analyze the parameters.

Parameters' chains and model fit

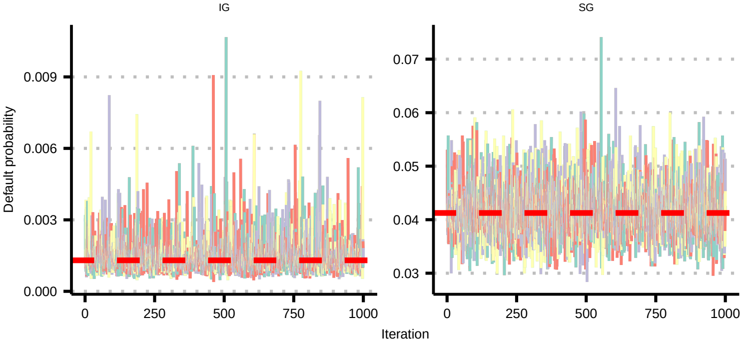

Figure 6: Default probability chains with mean value in red

Applying the model to S&P corporate default rates

Parameters' chains and model fit

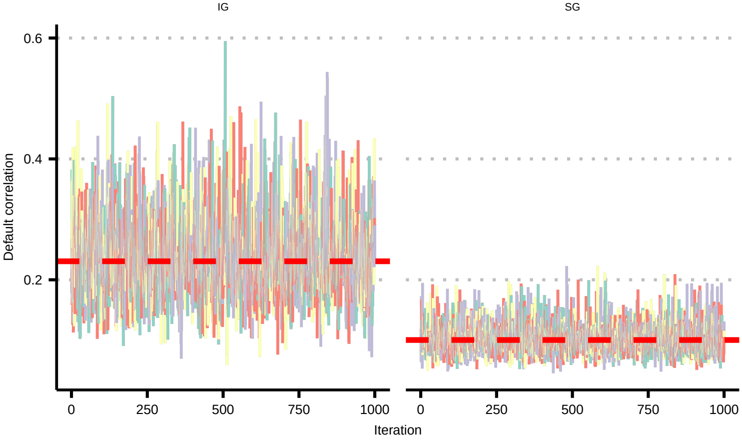

Figure 7: Default correlation chains with mean value in red

Applying the model to S&P corporate default rates

Parameters' chains and model fit

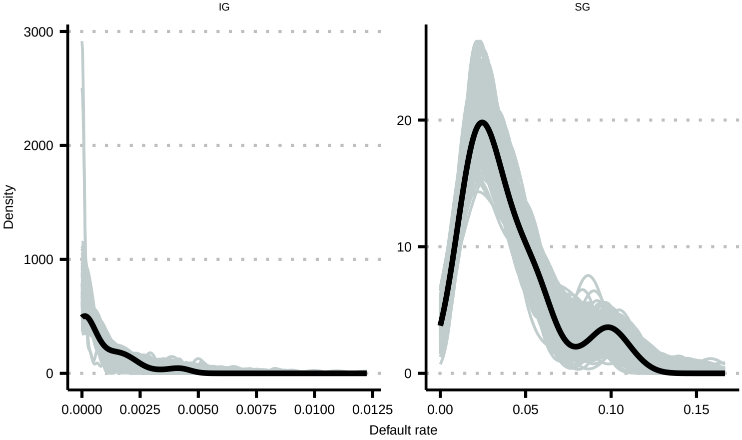

Figure 8: Density overlay for S&P corporate default rates

Applying the model to S&P corporate default rates

One step forecasts of the default rate

We perform forecasts of the form $D_{t+1} \mid D_1, .., D_t, N_{t+1}$ to assess the risk coverage of the model, using 50% and 90% posterior credibility intervals for D_{t+1}.

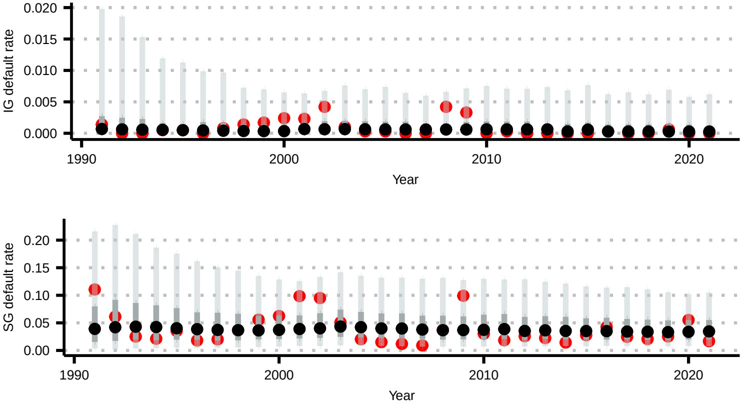

Figure 9: One step forecast intervals

Conclusion and future studies

The Bayesian framework can provide further insights regarding the estimates of the default probability and correlation. Furthermore, for small samples and low correlations, it can better capture the uncertainty regarding the model. Finally, the architecture surrounding Stan allows us to perform posterior predictive checks with ease and observe the goodness of fit to real data.

Possible advancements:

- Using AR(1) factors instead of normal
- Multivariate setting, estimating inter-group correlation with hierarchical models
- Using covariates for the factor regarding economic activity or financial stress