Similarly to (cite vasicek), we assume that the default of a loan happens when

$$A < B$$

where A is the value the debtor's (hypothetical) assets and B is the value of his debts. The recovery rate is, in line with (Pythkin 2003)k computed as

$$R = \frac{\min(P, p)}{p} = \min(p^{-1}P, 1)$$

where p is the outstanding principle of the loan and P is the price of the collateral.

As it is usual, we assume that

$$A = \exp\{X + Z\}, \qquad P = \exp\{H + E\}$$

where X, H are the factors, common for all the loans, and Z, E are mutually independent normally distributed individual factors, specific for each loan. For simplicity, we assume that B is common for all the loans.

If the loan portfolio is large and homogeneous, then the default rate (sometimes imprecisely referenced as probability of default - PD) of a large homogeneous loan portfolio may approximated as

$$Q = \frac{\text{number of defaults}}{\text{number of loans}} \doteq \mathbb{P}[A < B|Y] = \varphi(-Y), \qquad Y = \frac{X - \log B}{\rho}$$

where φ is a standard normal c.d.f. and ρ is the standard deviation of Z. The loss given default (LGD) - another quantity of usual interest - comes out as

$$G = \frac{\text{total loss of porfolio}}{\text{number of defaults}} \doteq \mathbb{E}[1 - R|I] = 1 - \mathbb{E}[\min(\exp\{I + E\}, 1) = \eta(I; \sigma)$$

$$I = H - \log p, \qquad \eta(\iota; \sigma) = e^{\iota} \int_{-\infty}^{-\iota} \varphi(\frac{x}{\rho}) e^{x} dx = \varphi(-\frac{\iota}{\sigma}) - \exp\{\iota + \frac{1}{2}\sigma^2\}\varphi(-\frac{\iota}{\sigma} - \sigma)$$

where σ is the standard deviation of E. For more detailed explanation, see either [GŠ FU 2012] or [Smid, Martin and Dufek, Jaroslav, Multi-Period Factor Model of a Loan Portfolio (July 10, 2016). Available at SSRN: http://dx.doi.org/10.2139/ssrn.2703884].

As for the dynamics, we suppose that the vector (X, B, H) follows a VAR model, i.e.

$$(X_t, B_t, H_t) = \Gamma U_t + \mathcal{E}_t$$

where Γ is unknown deterministic matrix parameter, $\mathcal{E}_t$ is a Gaussian white noise and where the regressors U_t are allowed to include constants, trends, lagged values and exogenous variables. Consequently,

$$Q_t = \varphi(-Y_t) = \varphi\left(-\left[\Gamma^Q U_i + \epsilon_{1,t}\right]\right),$$

$$G_t = \eta\left(I_t; \sigma\right) = \eta\left(\Gamma^G U_i + \varepsilon_{2,t}; \sigma\right)$$

for some vector parameters Γ^Q and Γ^G and a Gaussian white noise $(\epsilon_1, \varepsilon_2)$.

Once Q_t and G_t are observable, the factors might be got by transformation

$$Y_t = -\varphi^{-1}(Q_t), \qquad I_t = \eta^{-1}(G_t; \sigma)$$

where the (unknown) value of σ might be guessed e.g. from the volatility of a house price index and the average default rate - see the Appendix for details. Consequently, the parameters $\Gamma^\bullet$ and the variance matrix of ϵ may be estimated by standard techniques. For the proof of strict monotonicity of $\eta(\bullet; \sigma)$, see Appendix of [GŠ FÚ 2012].

If, instead of G_t, the charge-off rate

$$L_t = \frac{\text{total loss of porfolio}}{\text{number of loans}}$$

is observed, G_t may be easily computed using the relation $L_t = Q_t G_t$.

Appendix

Determination of σ

Assume that, at time t, the portfolio contains multiple "generations" of loans namely the loans originated at $t-1, t-2, \dots, t-k$ (the loans older than k are no longer present in the portfolio). Assume further that the inflow of fresh loans into the portfolio is constant in time. Finally, assume that all the collaterals securing loans from the generation which started at s have been bought for the same price $\exp\{H_s\}$ and that the price of each of them at t is $\exp\{H_t + (S_t - S_s)\}$ where S is a normal random walk, specific to the loan, with variance θ^2,

Denote G_t the age of a loan randomly chosen at t. Clearly, after k periods, the ratio of the generations within the portfolio is$1 : (1-q) : \dots : (1-q)^{k-1}$.which unequely determines $\pi_i = \mathbb{P}[G_t = i]$.

Let P_t be the price of a randomly chosen collateral. By the Law of Iterated Variance, we then get

$$\tilde{\sigma}^2 = \operatorname{var}(\log P_t | H) = \operatorname{var}(\mathbb{E}(\log P_t | G_t, H)|H) + \mathbb{E}(\operatorname{var}(\log P_t | G_t, H)|H)$$

$$= \operatorname{var}(\mathbb{E}(S_t - S_G | G, H)|H) + \mathbb{E}(\operatorname{var}(S_t - S_G | G, H)|H) = \theta^2 \, \mathbb{E}\boldsymbol{G_t} \boldsymbol{=} \theta^2 \sum_{i=1}^{k} i\pi_i,$$

Even though the $\mathcal{L}(\log P_t | H)$ is a mixture of normal distributions rather than a normal distribution, it is thin tailed so it will not make a big harm to approximate it by $\mathcal{N}(H_t, \tilde{\sigma}^2)$,