1 Methods

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) 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.

Analogously to the Vasicek model, we assume that

$$A = \exp\{Y^A + Z^A\}, \qquad B = \exp\{Y^B + Z^B\}, \qquad P = \exp\{X + E\}$$

where Y^A, Y^B, X are factors, common for all the loans, and Z^A, Z^B, E are individual factors, specific for each loan, such that (Z^A, Z^B) is Gaussian, independent of E which is normal.

If a loan portfolio is large and homogeneous then the default rate

$$Q = \frac{\text{number of defaults}}{\text{number of loans}}$$

may be, thanks to the Law of Large Numbers applied to conditional distributions given the common factors, approximated as

$$Q_t \doteq \mathbb{P}[A < B|Y^A, Y^B] = \mathbb{P}[Z^A - Z^B < Y^B - Y^A|Y^A, Y^B] = \varphi(-Y), \qquad Y = \frac{Y^A - Y^B}{\rho}$$

where φ is a standard normal c.d.f. and ρ is the standard deviation of $Z^A - Z^B$.

The loss given default (LGD) - another quantity of usual interest - is defined as

$$G = 1 - \frac{\text{total recovery from defaulted dents}}{\text{number of defaults}}$$

By Proposition 1 (i) and (ii),

$$G \doteq \mathbb{E}(1 - R|I) = h(I; \sigma)$$

$$I = X - \log p, \qquad h(\iota; \sigma) = \varphi(-\frac{\iota}{\sigma}) - \exp\{\iota + \frac{1}{2}\sigma^2\}\varphi(-\frac{\iota}{\sigma} - \sigma)$$

where σ is the standard deviation of E.

To introduce dynamics, we suppose that the vector (Y^A, Y^B, I) follows a VAR model, i.e.

$$(Y_t^A, Y_t^B, I_t) = \Gamma U_t + \mathcal{E}_t \tag{1}$$

where Γ is an (unknown) deterministic matrix, $\mathcal{E}_t$ is a Gaussian white noise and U_t is a matrix of regressors possibly including constants, trends, lagged values of Y^A, Y^B, I, their differences, and exogenous variables. Consequently,

$$Q_t = \varphi(-Y_t) = \varphi\left(-\left[\Gamma^Q U_t + \epsilon_{1,t}\right]\right), \tag{2}$$

$$G_t = h\left(I_t; \sigma\right) = h\left(\Gamma^G U_i + \varepsilon_{2,t}; \sigma\right) \tag{3}$$

for some vector parameters Γ^Q and Γ^G and a Gaussian white noise $(\epsilon_1, \varepsilon_2)$, i.e. the dynamics of Q_t and G_t is uniquely determined by that of Y_t and I_t. Moreover, thanks to strict monotonicity of φ and h (see Appendix of [GŠ FÚ 2012] for the latter), the correspondence between the losses and the factors is one-to-one.

Thus, if σ is known, (1), (2) and (5) may serve as a model for the dynamics of Q_t and G_t. In particular, once a time series of Q_t and $G_t$is observed, factors may be retrieved first by formula

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

and standard techniques may be used to estimate the parameters of the VAR model.

Making predictions in the model is rather straightforward. If history $\Omega_t = (U_\tau, Y_\tau, I_\tau)_{\tau \leq t}$ is observed and

$$Y_T | \Omega_t \sim \mathcal{N}(\mu, v^2)$$

for some $T > t$ and Ω_t-measurable μ and v, then

$$q_{t,T} = \mathbb{P}[A_T < B_T | \Omega_t]$$

i.e. the probability of default of a single loan given Ω_t comes out as

$$q_{t,T} = \mathbb{P}[Z_T^A - Z_T^B + Y_T^A - Y_T^B < 0 | \Omega_t] = \mathbb{P}\left[\frac{Z_T}{\rho} + Y_T \leq 0 \middle| \Omega_t\right] = \varphi\left(\frac{-\mu}{\sqrt{v^2+1}}\right)$$

while the distribution of the default rate is given by

$$\begin{aligned}
\mathbb{P}[Q_T < \theta | \Omega_t] &= \int \mathbf{1}\{\varphi(y) < \theta\} \mathrm{d}\mathbb{P}_{-Y_T|\Omega_t}(y) = \int \mathbf{1}\{\varphi(y) < \theta\} \mathrm{d}\varphi\left(\frac{y+\mu}{v}\right) \\
&= \int \mathbf{1}\{z < \theta\} \mathrm{d}\varphi\left(\frac{\varphi^{-1}(z)+\mu}{v}\right) = \varphi\left(\frac{\varphi^{-1}(\theta)+\mu}{v}\right)
\end{aligned}$$

with

$$\begin{aligned}
\mathbb{E}(Q_T | \Omega_t) &= \int z \mathrm{d}\varphi\left(\frac{\varphi^{-1}(z)+\mu}{v}\right) = \int \varphi^{-1}(x-\mu) \mathrm{d}\varphi\left(\frac{x}{v}\right) \\
&= \mathbb{P}[\mathcal{N}(0, v^2) - \mathcal{N}(0,1) \leq -\mu] = \varphi\left(\frac{-\mu}{\sqrt{v^2+1}}\right) = q_{t,T}
\end{aligned}$$

i.e. $q_{t,T}$ is a (conditionally) unbiased point forecast of Q_T. [1]

Analogously, if $I_T|\Omega_t \sim \mathcal{N}(\nu, w^2)$ then

$$g_{t,T} = \mathbb{E}[1 - R_T|\Omega_t]$$

- the mean loss of a loan given Ω_t - fulfills

$$g_t = h(\nu; \sqrt{\sigma^2 + w^2})$$

(see (iii) of Proposition 1) and

$$\mathbb{P}[G_T < \theta|\Omega_t] = \int \mathbf{1}\{h(\iota;\sigma) < \theta\} d\mathbb{P}_{I_T|\Omega_t}(\iota) = \int \mathbf{1}\{h(\iota;\sigma) < \theta\} d\varphi\left(\frac{\iota - \nu}{w}\right)$$

$$= \int \mathbf{1}\{z < \theta\}\varphi\left(\frac{h^{-1}(z;\sigma) - \nu}{w}\right) = \varphi\left(\frac{h^{-1}(\theta;\sigma) - \nu}{w}\right) = \varphi\left(\frac{h^{-1}(\theta;\sigma) - h^{\cdot 1}(g_{i,t}; \sqrt{\sigma + w^2})}{w}\right).$$

with

$$\mathbb{E}(G_T|\Omega_t) = \mathbb{E}(\mathbb{E}(1 - R_T|I_T)|\Omega_t) = \mathbb{E}(1 - R_T|\Omega_T) = g_{t,T}$$

(by the Chain Rule for Conditional Expectations).

Once $[Y_t^L, Y_t^H]$, $[I_t^L, I_t^H]$ are confidence intervals for future values of Y_t, I_t, the intervals $[\varphi(-Y_t^H), \varphi(-Y_t^L)]$, $[h(I_t^H), h(I_t^L)]$, are confidence intervals for Q_t, G_t, respectively.

Finally, the percentage loss of the portfolio

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

which is usually of primary interest, may be computed by formula

$$L_t = Q_t G_t = \varphi(-Y_t)h(I_t;\sigma).$$

Due to possible dependence of common factors forecasts, however, its conditional distribution, point forecasts and confidence intervals have to be computed by simulation.

[1] Note that, equivalently,

$$\mathbb{P}[Q_T < \theta|\Omega_t] = \varphi\left(\frac{1}{\sqrt{\vartheta}}\left(\sqrt{1-\vartheta}\varphi^{-1}(\theta) + \varphi^{-1}(q_{t,T})\right)\right)$$

where

$$\vartheta = \sqrt{\frac{v^2}{v^2+1}} = \mathrm{corr}((Y_T^A - Y_T^B) + Z_{T,1}, (Y_T^A - Y_T^B) + Z_{T,2}|\Omega), \qquad Z_{T,i} \sim \mathcal{N}(0,\rho), \quad i = 1,2, \qquad \perp\!\!\!\perp \Omega, Z_{T,1}, Z_{T,2}$$

i.e. our formula is an analog of the well known formula for the loss distribution in Vasicek model (1987).

2 Results

If σ is unknown then the 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 Γ^Q and Γ^G and the variance matrix of ϵ may be estimated by standard techniques.

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$1 : (1-q) : \cdots : (1-q)^{k-1}$.which uniquely 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

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

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

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 $\mathcal{N}(H_t, \tilde{\sigma}^2)$,

Properties of LGD

Proposition 1. *(i) With increasing size of portfolio,* $G \to \mathbb{E}(1-R|I)$
(ii) $\mathbb{E}(1-R|I) = h(J;\sigma)$
(iii) If $I = I_1 + I_2$, $I_1 \in \Omega$, $I_2|\Omega \sim \mathcal{N}(0, s^2)$, $E \perp\!\!\!\perp \Omega, I_2$ *for some sigma-field* Ω *then* $\mathbb{E}(h(I;\sigma)|\Omega) = h(I_1; \sqrt{s^2+\sigma^2})$

Proof. (i) TBD

(ii) and (iii): Let $\mathcal{F}$ be a sigma field and put $m(a,b) = \min(\exp\{a+b\}, 1)$. Let $A \in \mathcal{F}$, $B|\mathcal{F} \sim \mathcal{N}(0, s^2)$. Then

$$\mathbb{E}[m(A,B),1)|\mathcal{F}] = \int \min(e^{x+A}, 1) \mathrm{d}\varphi(\frac{x}{s}) = \int_{-\infty}^{-A} e^{x+A} \mathrm{d}\varphi(\frac{x}{\sigma}) + (1-\varphi(-\frac{A}{\sigma})) = 1 - h(A, s) \tag{4}$$

where the last equality follows from Appendix of FaU2012.
Using this and the fact that

$$R = \min(\exp\{I + E), 1\}) = m(I, E) \tag{5}$$

we are getting (ii). As for (iii), note first that

$$I, E \perp\!\!\!\perp_{I_1} \Omega \tag{6}$$

(to see it, note that, by the Chain Rule for independence $I_2, E \perp\!\!\!\perp \Omega$ trivially implying $I_2, E \perp\!\!\!\perp \Omega, I_1$, giving, by Kallenberg Proposition 6.8, that $I_2, E \perp\!\!\!\perp_{I_1} \Omega$, giving (6) by Corollary 6.7. (i)). By using (ii), (5), (6) the Law of Iterated Expectation and (4), we have

$$\begin{aligned}
\mathbb{E}(1 - h(I; \sigma)|\Omega) &= \mathbb{E}(\mathbb{E}(R|I)|\Omega) = \mathbb{E}(\mathbb{E}(m(I, E)|I)|\Omega) \\
&= \mathbb{E}(\int m(x, y) d\mathbb{P}_{I,E}(x, y|I)|\Omega) = \mathbb{E}(\int m(x, y) d\mathbb{P}_{I,E}(x, y|I, \Omega)|\Omega)
\end{aligned}$$
$$= \mathbb{E}(\mathbb{E}(m(I, E)|I, \Omega)|\Omega) = \mathbb{E}(\mathbb{E}(m(I, E)|\Omega) = \mathbb{E}(m(I_1, I_2 + E)|\Omega) = 1 - h(I_1, \sqrt{s^2 + \sigma^2}).$$

TBD možná přes integrály. □