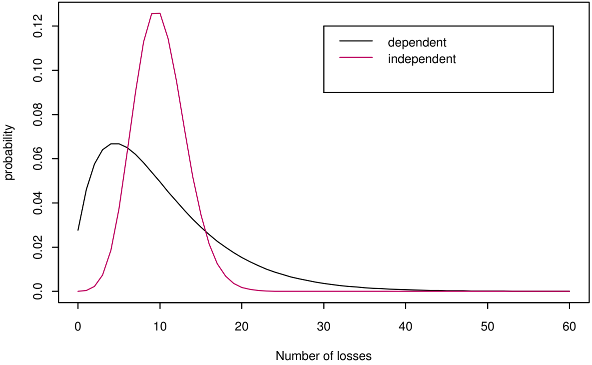
11 Portfolio credit risk management

Importance of default dependence

Dependence between defaults (and downgrades) is a key issue in credit risk management. There are two main sources of dependence between defaults:

- Dependence caused by common factors (for example, interest rates and changes in economic growth) affecting all obligors
- Default of company A may have direct impact on default probability of company B and vice versa because of direct business relations, a phenomenon known as contagion



Comparison of the loss distribution of a homogeneous portfolio of 1000 loans with a default probability of $p_1 = \cdots = p_{1000} = 1\%$ assuming (i) independent defaults and (ii) a default correlation of $\rho(Y_i, Y_j) = 0.5\%$. Case (ii) can be considered as roughly representative for BB-rated loans.

11.1 Threshold models

11.1.1 Notation for one-period portfolio models

- Consider portfolio of m firms and time horizon $T = 1$ (say one year).
- For $1 \leq i \leq m$, let R_i be a state indicator for obligor i at time T taking values in the set $\{0, 1, \ldots, n\}$; we interpret the value 0 as default and non-zero values as states of increasing credit quality. At time $t = 0$ obligors are assumed to be in some non-default state.
- Mostly we will concentrate on the binary outcomes of default and non-default. We write Y_i for the default indicator variables so that $Y_i = 1 \iff R_i = 0$ and $Y_i = 0 \iff R_i > 0$.
- The random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_m)'$ is a vector of default indicators for the portfolio and $p(\boldsymbol{y}) = \mathbb{P}(Y_1 = y_1, \ldots, Y_m = y_m)$, $\boldsymbol{y} \in \{0, 1\}^m$, is its joint probability function; the marginal default probabilities are denoted by $p_i = \mathbb{P}(Y_i = 1)$, $i = 1, \ldots, m$.

- Default or event correlation. Noting that

$$\operatorname{var}(Y_i) = \mathbb{E}(Y_i^2) - p_i^2 = \mathbb{E}(Y_i) - p_i^2 = p_i - p_i^2,$$

we obtain, for firms i and j with $i \neq j$, the formula

$$\rho(Y_i, Y_j) = \frac{\mathbb{E}(Y_i Y_j) - p_i p_j}{\sqrt{(p_i - p_i^2)(p_j - p_j^2)}}. \tag{111}$$

- Let the rv $M := \sum_{i=1}^m Y_i$ denote the number of defaulted obligors at T.
- The actual loss if company i defaults is modelled by the random quantity $\delta_i e_i$, where e_i represents the overall exposure to company i and $0 \leq \delta_i \leq 1$ represents the LGD.
- We denote the overall portfolio loss by $L := \sum_{i=1}^m \delta_i e_i Y_i$.
- It is possible to set up different credit risk models leading to the same multivariate distribution for $\boldsymbol{R}$ or $\boldsymbol{Y}$. We call two models with state vectors $\boldsymbol{R}$ and $\tilde{\boldsymbol{R}}$ (or $\boldsymbol{Y}$ and $\tilde{\boldsymbol{Y}}$) equivalent if $\boldsymbol{R} \stackrel{\mathrm{d}}{=} \tilde{\boldsymbol{R}}$ (or $\boldsymbol{Y} \stackrel{\mathrm{d}}{=} \tilde{\boldsymbol{Y}}$).

The exchangeable special case

- It is common to group obligors together to form homogeneous groups. This corresponds to the mathematical concept of exchangeability.
- A random vector $\mathbf{R}$ is exchangeable if

$$(R_1, \ldots, R_m) \stackrel{\text{d}}{=} \left(R_{\Pi(1)}, \ldots, R_{\Pi(m)}\right),$$

  for any permutation $(\Pi(1), \ldots, \Pi(m))$ of $(1, \ldots, m)$.
- We talk of an exchangeable default model if the default indicator vector $\mathbf{Y} = (Y_1, \ldots, Y_m)'$ is exchangeable.
- Note that this permits a simple notation for default probabilities:

$$\pi_k := \mathbb{P}\left(Y_{i_1} = 1, \ldots, Y_{i_k} = 1\right), \quad \{i_1, \ldots, i_k\} \subset \{1, \ldots, m\},$$
$$\pi := \pi_1 = P\left(Y_i = 1\right), \quad i \in \{1, \ldots, m\}.$$

- In the exchangeable case the default correlation is given by

$$\rho_Y := \rho(Y_i, Y_j) = \frac{\pi_2 - \pi^2}{\pi - \pi^2}, \quad i \neq j. \tag{112}$$

11.1.2 Threshold models and copulas

Definition 11.1

Let $\boldsymbol{X} = (X_1, \dots, X_m)'$ be an m-dimensional random vector and let $D \in \mathbb{R}^{m \times n}$ be a deterministic matrix with elements d_{ij} such that, for every i, the elements of the ith row form a set of increasing thresholds satisfying $d_{i1} < \cdots < d_{in}$. Augment these thresholds by setting $d_{i0} = -\infty$ and $d_{i(n+1)} = \infty$ for all obligors and then set

$$R_i = j \iff d_{ij} < X_i \le d_{i(j+1)}, \quad j \in \{0, \dots, n\},\ i \in \{1, \dots, m\}.$$

Then $(\boldsymbol{X}, D)$ is said to define a threshold model for $\boldsymbol{R} = (R_1, \dots, R_m)'$.

- $\boldsymbol{X}$ are the critical variables and the ith row of D contains the critical thresholds for firm i.
- Default occurs if $X_i \le d_{i1}$ so that the default probability of company i is given by $p_i = F_{X_i}(d_{i1})$.

- When working with a default-only model we simply write $d_i = d_{i1}$ and denote the threshold model by $(\boldsymbol{X}, \boldsymbol{d})$.
- Default correlation and asset correlation. It is important to distinguish the default correlation $\rho(Y_i, Y_j)$ of two firms $i \neq j$ from the correlation of the critical variables X_i and X_j.
- Since the critical variables are often interpreted in terms of asset values, the latter correlation is often referred to as asset correlation.
- For given default probabilities, $\rho(Y_i, Y_j)$ is determined by $\mathbb{E}(Y_i Y_j)$ according to (111), and in a threshold model $\mathbb{E}(Y_i Y_j) = \mathbb{P}(X_i \leq d_{i1},\ X_j \leq d_{j1})$, so default correlation depends on the joint df of X_i and X_j.
- If $\boldsymbol{X}$ is multivariate normal, as in many models used in practice, the correlation of X_i and X_j determines the copula of their joint distribution and hence the default correlation.

Copulas: Key facts for this chapter

- A copula is a multivariate distribution function (df) with standard uniform marginal distributions (margins).
- If the df of $\boldsymbol{U} = (U_1, \ldots, U_m)'$ is the copula C, we have that $\mathbb{P}(U_i \leq u_i) = C(1, \ldots, 1, u_i, 1 \ldots, 1) = u_i$.

Sklar's Theorem

Let F be a joint distribution function with margins $F_1, \ldots, F_m$. There exists a copula C such that for all $x_1, \ldots, x_m$ in $[-\infty, \infty]$

$$F(x_1, \ldots, x_m) = C(F_1(x_1), \ldots, F_m(x_m)).$$

If the margins are continuous then C is unique. Conversely, if C is a copula and $F_1, \ldots, F_m$ are univariate distribution functions, then F defined above is a multivariate df with margins $F_1, \ldots, F_m$.

- Sklar's Theorem shows that every multivariate df can be written as a copula function of its marginal dfs.
- The converse shows that copulas can be used to create multivariate distributions with arbitrary margins.
- Let $\boldsymbol{X} = (X_1, \ldots, X_m)'$ be a random vector with df F and continuous margins.
  - We refer to the copula C contained in F as the copula of F (or $\boldsymbol{X}$).
  - Invariance property: C is also the copula of $(T_1(X_1), \ldots, T_m(X_m))$ for strictly increasing transformations $T_1, \ldots, T_m$.
  - The copula C can be viewed as a representation of the dependence structure of F (or $\boldsymbol{X}$).
- The copula of a multivariate normal vector $\boldsymbol{X}$ is known as the Gauss copula. In view of the invariance property it only depends on the correlation matrix $\mathbb{P}$ of $\boldsymbol{X}$.

- The copula of a multivariate Student t random vector $\boldsymbol{X}$ is known as the t copula. It depends on the correlation matrix of $\boldsymbol{X}$ and the degree of freedom ν.
- The Gauss and t copula do not have simple analytical forms. They are, however, flexible dependence models because they have at least one parameter for every pairs of marginal distributions.
- There are other copulas with simpler forms like the Archimedean copulas. These take the form

$$C(\boldsymbol{u}) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_m)), \quad \boldsymbol{u} \in [0,1]^m,$$

  where the *(Archimedean) generator* $\psi : [0,\infty) \to [0,1]$ is strictly decreasing on $[0, \inf\{t : \psi(t) = 0\}]$ and satisfies $\psi(0) = 1$ and $\lim_{t\to\infty} \psi(t) = 0$. Examples include Clayton ($\psi(t) = (1+t)^{-1/\theta}$, $\theta \in (0,\infty)$) and Gumbel ($\psi(t) = \exp(-t^{1/\theta})$, $\theta \in [1,\infty)$). These are distributions for exchangeable random vectors $\boldsymbol{U}$.

Copulas in threshold models

Lemma 11.2

Let $(\boldsymbol{X}, D)$ and $(\tilde{\boldsymbol{X}}, \tilde{D})$ be a pair of threshold models with state vectors $\boldsymbol{R} = (R_1, \ldots, R_m)'$ and $\tilde{\boldsymbol{R}} = (\tilde{R}_1, \ldots, \tilde{R}_m)'$, respectively. The models are equivalent if the following conditions hold.

(i) The marginal distributions of the random vectors $\boldsymbol{R}$ and $\tilde{\boldsymbol{R}}$ coincide, i.e.

$$\mathbb{P}(R_i = j) = \mathbb{P}(\tilde{R}_i = j), \quad j \in \{1, \ldots, n\},\ i \in \{1, \ldots, m\}.$$

(ii) $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$ admit the same copula C.

- The copula is critical for joint default probabilities. Consider a subgroup of k companies $\{i_1, \ldots, i_k\} \subset \{1, \ldots, m\}$.

- We have

$$\begin{aligned}\mathbb{P}(Y_{i_1} = 1, \ldots, Y_{i_k} = 1) &= \mathbb{P}\left(X_{i_1} \le d_{i_1}, \ldots, X_{i_k} \le d_{i_k}\right) \\ &= C_{i_1,\ldots,i_k}\left(F_{i_1}(d_{i_1}), \ldots, F_{i_k}(d_{i_k})\right) \\ &= C_{i_1,\ldots,i_k}(p_{i_1}, \ldots, p_{i_k})\,,\end{aligned}$$

  where $C_{i_1,\ldots,i_k}$ is a k-dimensional marginal df of C.
- For $S \subset \{1, \ldots, d\}$, C_S is obtained from $C(u_1, \ldots, u_d)$ by setting $u_i = 1$ for $i \notin S$.
- The copula C crucially determines higher order joint default probabilities and thus extreme risk that many companies default.
- In an exchangeable default model we have

$$\pi_k = C_{1,\ldots,k}(\pi, \ldots, \pi), \quad 2 \le k \le m.$$

11.1.3 Gaussian threshold models

Multivariate Merton model:

- Assume that the multivariate asset-value process $\boldsymbol{V}_t = (V_{t,1}, \dots, V_{t,m})'$ follows an m-dimensional GBM with drift vector $\boldsymbol{\mu}_V = (\mu_1, \dots, \mu_m)'$, vector of volatilities $\boldsymbol{\sigma}_V = (\sigma_1, \dots, \sigma_m)'$ and correlation matrix P.
- This means that $(\boldsymbol{V}_t)$ solves the stochastic differential equations

$$\mathrm{d}V_{t,i} = \mu_i V_{t,i}\,\mathrm{d}t + \sigma_i V_{t,i}\,\mathrm{d}W_{t,i}, \quad i = 1, \dots, m,$$

  for correlated BMs with correlation $\rho(W_{t,i}, W_{t,j}) = \rho_{ij}$, $t \geq 0$.
- For all i the asset value $V_{T,i}$ is of the form

$$V_{T,i} = V_{0,i} \exp\big((\mu_i - \tfrac{1}{2}\sigma_i^2)T + \sigma_i W_{T,i}\big),$$

  where $\boldsymbol{W}_T \sim \mathrm{N}_m(\boldsymbol{0}, TP)$.
- In its basic form the Merton model is a default-only model where the firm defaults if $V_{T,i} \leq B_i$ and B_i is the liability of firm i.

- Writing $\boldsymbol{B} = (B_1, \ldots, B_m)'$ the threshold model representation is $(\boldsymbol{V}_T, \boldsymbol{B})$.
- The multivariate Merton model is equivalent to the model $(\boldsymbol{X}, \boldsymbol{d})$ with

$$X_i := \frac{\ln V_{T,i} - \ln V_{0,i} - (\mu_i - \frac{1}{2}\sigma_i^2)T}{\sigma_i\sqrt{T}},$$
$$d_i := \frac{\ln B_i - \ln V_{0,i} - (\mu_i - \frac{1}{2}\sigma_i^2)T}{\sigma_i\sqrt{T}}.$$

- The transformed variables satisfy $\boldsymbol{X} \sim \mathrm{N}_m(\mathbf{0}, P)$ and their copula is the Gauss copula C_P^{Ga}.

Gaussian threshold models in practice

- In practice it is usual to start directly with threshold models of the form $(\boldsymbol{X}, \boldsymbol{d})$ with $\boldsymbol{X} \sim \mathrm{N}_m(\mathbf{0}, P)$.

- There are two practical challenges:
  1) calibration of $\boldsymbol{d}$ (or, in the case of a multi-state model, the threshold matrix D) in line with exogenously given default and transition probabilities;
  2) calibration of the correlation matrix P in a parsimonious way.
- The problem of embedding state transition probabilities in a threshold matrix D has already been discussed. In a default-only model we set $d_i = \Phi^{-1}(p_i)$ for $i = 1, \dots, m$.

Factor models

- In its most general form P has $m(m-1)/2$ distinct parameters.
- m is typically large and it is important to use a more parsimonious parametrization of this matrix based on a factor model.
- Factor models also lend themselves to economic interpretation and the factors are commonly interpreted as country and industry effects.

- We assume that
$$X_i = \sqrt{\beta_i}\tilde{F}_i + \sqrt{1-\beta_i}\varepsilon_i, \tag{113}$$
where $\tilde{F}_i$ and $\varepsilon_1, \dots, \varepsilon_m$ are independent standard normal variables, and where $0 \le \beta_i \le 1$ for all i.
- In this formulation $\tilde{F}_i$ are the systematic variables, which are correlated, and ε_i are idiosyncratic variables.
- It follows that β_i can be viewed as a measure of the systematic risk of X_i: that is, the part of the variance of X_i which is explained by the systematic variable.
- The systematic variables are assumed to be of the form $\tilde{F}_i = \boldsymbol{a}_i'\boldsymbol{F}$ where $\boldsymbol{F}$ is a vector of common factors satisfying $\boldsymbol{F} \sim \mathrm{N}_p(\mathbf{0}, \Omega)$ with $p < m$, and where Ω is a correlation matrix.
- These factors typically represent country and industry effects.
- The assumption that $\operatorname{var}(\tilde{F}_i) = 1$ means that $\boldsymbol{a}_i'\Omega\boldsymbol{a}_i = 1$ for all i.

- Since $\operatorname{var}(X_i) = 1$ and since $\tilde{F}_i$ and $\varepsilon_1, \ldots, \varepsilon_m$ are independent and standard normal, the asset correlations in this model are given by

$$\rho(X_i, X_j) = \operatorname{cov}(X_i, X_j) = \sqrt{\beta_i \beta_j}\operatorname{cov}(\tilde{F}_i, \tilde{F}_j) = \sqrt{\beta_i \beta_j}\boldsymbol{a}_i' \Omega \boldsymbol{a}_j \,.$$

- In order to set up the model we have to determine $\boldsymbol{a}_i$ and β_i for each obligor and Ω, with the additional constraint that $\boldsymbol{a}_i' \Omega \boldsymbol{a}_i = 1$ for all i.
- Since Ω has $p(p-1)/2$ parameters, the loading vectors $\boldsymbol{a}_i$ and coefficients β_i have collectively $mp + m$ parameters, and we are applying m constraints, this gives $mp + p(p-1)/2$ parameters.

The one-factor model

- We often consider the special case of a one-factor model.
- This corresponds to a model where $\tilde{F}_i = F$ for a single common standard normal factor so that the equation in (113) takes the form

$$X_i = \sqrt{\beta_i} F + \sqrt{1 - \beta_i}\varepsilon_i \,. \tag{114}$$

- If, moreover, every obligor has the same systematic variance $\beta_i = \rho$ we get that $\rho(X_i, X_j) = \rho$ for all $i \neq j$, which is often referred to as an equicorrelation model.

11.1.4 Models based on alternative copulas

t copula model

- Suppose Z_i follows the Gaussian factor model

$$Z_i = \sqrt{\beta_i}\tilde{F}_i + \sqrt{1-\beta_i}\varepsilon_i,$$

  with $\tilde{F}_i = \boldsymbol{a}_i'\boldsymbol{F}$ and all assumptions as before.

- Let $X_i = \sqrt{W} Z_i$ for $i = 1, \ldots, m$, where W has an inverse gamma distribution, $W \sim \mathrm{IG}(\frac{1}{2}\nu, \frac{1}{2}\nu)$, or equivalently, $\nu/W \sim \chi^2_\nu$.
- The vector $(X_1, \ldots, X_m)$ has a multivariate t distribution with ν degrees of freedom, location vector $\mathbf{0}$ and correlation matrix P identical to that of $(Z_1, \ldots, Z_m)$.

- The critical variables have the t copula $C^{\mathrm{t}}_{\nu,P}$.
- The class of threshold models based on the t copula can be thought of as containing the Gaussian threshold models as limiting cases when $\nu \to \infty$. However, the additional parameter ν adds a great deal of flexibility.

Archimedean copulas

- We recall that these take the form

$$C(u_1, \dots, u_m) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_m)),$$

  where the generator $\psi : [0, \infty) \to [0, 1]$ is a continuous, decreasing function satisfying $\psi(0) = 1$ and $\lim_{t\to\infty} \psi(t) = 0$, and ψ^{-1} is its inverse.
- The Clayton copula has generator $\psi_\theta(t) = (1 + \theta t)^{-1/\theta}$, where $\theta > 0$, leading to the expression

$$C^{\mathrm{Cl}}_\theta(u_1, \dots, u_m) = (u_1^{-\theta} + \cdots + u_m^{-\theta} + 1 - m)^{-1/\theta}.$$

- Suppose that $\boldsymbol{X}$ is a random vector with an Archimedean copula and marginal distributions F_{X_i}, $1 \leq i \leq m$, so that $(\boldsymbol{X}, \boldsymbol{d})$ specifies a threshold model with individual default probabilities $F_{X_i}(d_i)$.
- Consider the Clayton copula and assume a homogeneous situation where all individual default probabilities are identical to π.
- We can calculate that

$$\pi_k = (k\pi^{-\theta} - k + 1)^{-1/\theta}.$$

- Essentially, the dependent default mechanism of the homogeneous group is now determined by this equation and the parameters π and θ.

11.2 Mixture models

11.2.1 Bernoulli mixture models

- In a mixture model the default risk of an obligor is assumed to depend on a set of common factors, usually interpreted as macroeconomic variables, which are also modelled stochastically.
- Given a realization of the factors, defaults of individual firms are assumed to be independent.
- Dependence between defaults stems from the dependence of individual default probabilities on the set of common factors.
- Bernoulli mixture models provide a way of capturing the dependence between Bernoulli events (i.e. defaults/non-defaults).
- They can be extended to multinomial mixture models to capture dependent migrations in a rating system.

Definition 11.3 (Bernoulli mixture model)
Given some $p < m$ and a p-dimensional random vector $\boldsymbol{\Psi} = (\Psi_1, \dots, \Psi_p)'$, the default indicator vector $\mathbf{Y}$ follows a Bernoulli mixture model with factor vector $\boldsymbol{\Psi}$ if there are functions $p_i : \mathbb{R}^p \to (0,1)$, such that conditional on $\boldsymbol{\Psi}$ the components of $\mathbf{Y}$ are independent Bernoulli rvs with $\mathbb{P}(Y_i = 1 \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) = p_i(\boldsymbol{\psi})$.

The conditional independence given factors makes these models relatively easy to analyse. For $\mathbf{y} = (y_1, \dots, y_m)'$ in $\{0,1\}^m$ we get

$$\mathbb{P}(\mathbf{Y} = \mathbf{y} \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) = \prod_{i=1}^{m} p_i(\boldsymbol{\psi})^{y_i} (1 - p_i(\boldsymbol{\psi}))^{1-y_i}$$

$$\mathbb{P}(\mathbf{Y} = \mathbf{y}) = \int_{\mathbb{R}^p} \prod_{i=1}^{m} p_i(\boldsymbol{\psi})^{y_i} (1 - p_i(\boldsymbol{\psi}))^{1-y_i} g(\boldsymbol{\psi}) \mathrm{d}\boldsymbol{\psi},$$

where $g(\boldsymbol{\psi})$ is the probability density of the factors. The default probabilities are given by $p_i = \mathbb{E}(Y_i = 1) = \mathbb{E}(p_i(\boldsymbol{\Psi}))$.

- Consider the portfolio loss $L = \sum_{i=1}^{m} e_i \delta_i Y_i$ in the case where the exposures e_i and LGDs δ_i are deterministic.
- It is difficult to compute the df F_L of L.
- However, it is easy to use the conditional independence of the defaults to show that the *Laplace–Stieltjes transform* of F_L is for $t \in \mathbb{R}$ given by

$$\begin{aligned}
\hat{F}_L(t) = \mathbb{E}(e^{-tL}) &= \mathbb{E}\Big(\mathbb{E}(e^{-t\sum_{i=1}^{m} e_i\delta_i Y_i} \mid \boldsymbol{\Psi})\Big) \\
&= \mathbb{E}\bigg(\prod_{i=1}^{m} \mathbb{E}(e^{-te_i\delta_i Y_i} \mid \boldsymbol{\Psi})\bigg) \\
&= \mathbb{E}\bigg(\prod_{i=1}^{m} \big(p_i(\boldsymbol{\Psi})e^{-te_i\delta_i} + 1 - p_i(\boldsymbol{\Psi})\big)\bigg)
\end{aligned}$$

  which can be obtained by integrating over distribution of factors $\boldsymbol{\Psi}$.
- This is useful for: sampling losses from model with importance sampling; approximating probability mass function using Fourier inversion.

11.2.2 One-factor Bernoulli mixture models

Often it is useful to work with a one factor model ($p = 1$):

- Fitting to default data is relatively easy because evaluation of joint distribution/likelihood involves a one-dimensional integral at worst.
- The behaviour of large portfolios is easy to understand in terms of the distribution of the common factor.
- A one-factor model underlies the Basel II formula for computing capital under the IRB approach.

Thus we consider a rv Ψ and functions $p_i(\Psi)$ such that, conditional on Ψ, the default indicator vector $\boldsymbol{Y}$ is a vector of independent Bernoulli random variables with

$$\mathbb{P}(Y_i = 1 \mid \Psi = \psi) = p_i(\psi).$$

Exchangeable special case

- In the exchangeable special case of a one-factor model the conditional default probabilities $p_i(\Psi)$ are identical, $\forall i$, making $\mathbf{Y}$ exchangeable.
- In an exchangeable model we will introduce a new rv $Q = p_1(\Psi)$ with df $G(q)$ for the conditional default probability.
- Obviously Q has a distribution on $[0, 1]$.
- Recalling the π_k notation for exchangeable models we can calculate that

$$\pi := \mathbb{P}\left(Y_i = 1\right) = \mathbb{E}\left(Y_i\right) = \mathbb{E}\left(\mathbb{E}\left(Y_i \mid Q\right)\right) = \mathbb{E}(Q)$$

$$\pi_k := \mathbb{P}\left(Y_{i_1} = 1, \ldots, Y_{i_k} = 1\right) = \mathbb{E}(Q^k) = \int_0^1 q^k dG(q)\,. \tag{115}$$

- Conditional on $Q = q$, the number of defaults M is the sum of m independent Bernoulli variables and thus has a binomial distribution. The unconditional distribution of M is

$$\mathbb{P}(M = k) = \binom{m}{k} \int_0^1 q^k (1-q)^{m-k}\, \mathrm{d}G(q). \tag{116}$$

- In an exchangeable model the default probability and the higher order joint default probabilities are moments of the distribution of Q.
- Recall that the default correlation between two firms $i \neq j$ is defined to be the correlation between the default indicators Y_i and Y_j.
- In exchangeable Bernoulli mixtures we have

$$\operatorname{cov}(Y_i, Y_j) = \pi_2 - \pi^2 = \operatorname{var}(Q) \geq 0.$$

- The default correlation is given by

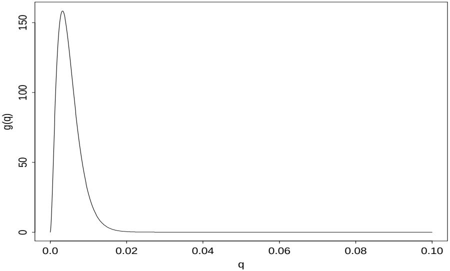
$$\rho_Y := \operatorname{corr}(Y_i, Y_j) = \frac{\pi_2 - \pi^2}{\pi - \pi^2} = \frac{\operatorname{var}(Q)}{\mathbb{E}(Q) - \mathbb{E}(Q)^2}.$$

Common mixing distributions

- Beta. $Q \sim \text{Beta}(a, b)$ where $a, b > 0$. This model corresponds quite closely to an exchangeable version of an industry model called CreditRisk$^+$ which we study later.

- Probit-Normal. $Q = \Phi(\mu + \sigma\Psi)$, $\Psi \sim \mathrm{N}(0,1)$
  This corresponds to an exchangable version of the threshold model with a Gaussian copula.
- Logit-Normal. $Q = (1 + \exp(-\mu - \sigma\Psi))^{-1}$, $\Psi \sim \mathrm{N}(0,1)$

In all these 2-parameter examples, if we fix default probability π and default correlation ρ_Y (or π_2) we fully calibrate the model. Picture shows beta density $g(q)$ of mixing variable Q in exchangeable Bernoulli mixture model with $\pi = 0.005$ and $\rho_Y = 0.0018$.



Example 11.4 (Beta mixing distribution)

- The density of a beta distribution is given by

$$g(q) = \frac{1}{\beta(a,b)} q^{a-1}(1-q)^{b-1}, \quad a,b > 0,\ 0 < q < 1,$$

  where $\beta(a,b)$ denotes the beta function.

- The beta function satisfies the following recursion formula: $\beta(a+1,b) = (a/(a+b))\beta(a,b)$.

- Using (115) we obtain for the higher-order default probabilities

$$\pi_k = \frac{1}{\beta(a,b)} \int_0^1 q^k q^{a-1}(1-q)^{b-1}\,\mathrm{d}q = \frac{\beta(a+k,b)}{\beta(a,b)}, \quad k = 1,2,\ldots.$$

- The recursion formula for the beta function yields: $\pi_k = \prod_{j=0}^{k-1}(a+j)/(a+b+j)$.

- In particular, $\pi = a/(a+b)$, $\pi_2 = \pi(a+1)/(a+b+1)$ and $\rho_Y = (a+b+1)^{-1}$.

- The rv M in (116) has a so-called *beta-binomial distribution*:

$$\mathbb{P}(M=k) = \binom{m}{k} \frac{1}{\beta(a,b)} \int_0^1 q^{k+a-1}(1-q)^{m-k+b-1}\,\mathrm{d}q$$
$$= \binom{m}{k} \frac{\beta(a+k, b+m-k)}{\beta(a,b)}. \tag{117}$$

One-factor models with covariates

- It is straightforward to extend the one-factor probit-normal and logit-normal mixture models to include covariates.
- These might be indicators for group membership, such as rating class or industry sector, or key ratios taken from a company's balance sheet.

- Writing $\boldsymbol{x}_i \in \mathbb{R}^k$ for a vector of covariates, we assume that

$$\begin{aligned} p_i(\Psi) &= h(\mu_i + \sigma_i \Psi) \\ \mu_i &= \mu + \boldsymbol{\beta}' \boldsymbol{x}_i \\ \sigma_i &= \exp(\delta + \boldsymbol{\gamma}' \boldsymbol{x}_i) \end{aligned} \tag{118}$$

  where $\Psi \sim \mathrm{N}(0,1)$, $h(x) = \Phi(x)$ or $h(x) = (1 + \exp(-x))^{-1}$.

- The vectors $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_k)'$ contain regression parameters, and $\mu \in \mathbb{R}$ and $\delta \in \mathbb{R}$ are intercept parameters.
- Similar specifications are commonly used in the class of generalized linear models in statistics.
- The regression structure in (118) includes partially exchangeable models where we define a number of groups within which risks are exchangeable. These groups might represent rating classes.

- If the covariates $\boldsymbol{x}_i$ are of the form $\boldsymbol{x}_i = \boldsymbol{e}_{r(i)}$, where $r(i) \in \{1, \ldots, k\}$ indicates the rating class of firm i, then the model (118) can be written in the form

$$p_i(\Psi) = h(\mu_{r(i)} + \sigma_{r(i)}\Psi) \tag{119}$$

for parameters $\mu_r := \mu + \beta_r$ and $\sigma_r := \exp(\delta + \gamma_r)$ for $r = 1, \ldots, k$.

- Suppose there are m_r obligors in rating category r for $r = 1, \ldots, k$, and write M_r for the number of defaults.

- The conditional distribution of the vector $\boldsymbol{M} = (M_1, \ldots, M_k)'$ is

$$\mathbb{P}(\boldsymbol{M} = \boldsymbol{l} \mid \Psi = \psi) = \prod_{r=1}^{k} \binom{m_r}{l_r} (h(\mu_r + \sigma_r\psi))^{l_r} (1 - h(\mu_r + \sigma_r\psi))^{m_r - l_r},$$

where $\boldsymbol{l} = (l_1, \ldots, l_k)'$.

11.2.3 Recovery risk in mixture

- In standard portfolio risk models it is assumed that the loss given default is independent of the default event.

- However, economic intuition suggests that recovery rates depend on similar risk factors to default probabilities.
- For example, during a property crisis many mortgages default. At the same time property prices are low, so that real estate can be sold only for very low prices in a foreclosure, leading to low recovery rates.
- The presence of systematic recovery risk is confirmed in a number of empirical studies, including Frye (2000) and Hamilton et al. (2005).
- The latter estimated that the relationship between the one-year default rate q and recovery rate R for corporate bonds was $R(q) \approx (0.52 - 6.9q)^+$.
- Systematics recoveries (LGDs) can be incorporated in the mixture-model framework by we replacing the constant δ_i with some function $\delta_i(\boldsymbol{\psi})$.
- The challenge lies in calibrating the function $\delta_i(\cdot)$.

11.2.4 Threshold models as mixture models

- Although the mixture models of this section seem, at first glance, to be different in structure to the threshold models, it is important to realize that the majority of useful threshold models, including all the examples we have given, can be represented as Bernoulli mixture models.
- Recall the threshold models based on a vector of multivariate Gaussian critical variables. These models can be motivated by a multivariate firm value model. Industry models such as Moody's public-firm EDF model or CreditMetrics belong in this category.
- Default occurs for counterparty i if a critical variable X_i lies below a critical threshold d_i. $\mathbf{X} = (X_1, \ldots, X_m)'$ is a random vector with standard normal margins (since we can standardize X_i and d_i without altering the default probability.)

- Moreover X_i follows a linear factor model

$$X_i = \sqrt{\beta_i}\boldsymbol{a}_i'\boldsymbol{F} + \sqrt{1-\beta_i}\varepsilon_i$$

where

  - $\boldsymbol{F} \sim \mathrm{N}_p(\mathbf{0}, \Omega)$ is a random vector of normally distributed common economic factors;
  - $0 \le \beta_i \le 1$ and $\mathrm{var}(\boldsymbol{a}_i'\mathbf{F}) = 1$;
  - $\varepsilon_1, \ldots, \varepsilon_m$ are iid standard normal and are also independent of $\boldsymbol{F}$.

- We will write the Gaussian threshold model as a Bernoulli mixture model with factor vector $\boldsymbol{\Psi} = -\boldsymbol{F}$. (This makes the conditional default probabilities increasing in the factors for positive $\boldsymbol{a}_i$.)
- Conditioning on $\boldsymbol{\Psi} = -\boldsymbol{F}$, the vector $\boldsymbol{X}$ is multivariate normally distributed with a diagonal covariance matrix and therefore the components of $\boldsymbol{X}$ are conditionally independent.

- The conditional default probabilities are

$$\begin{aligned} p_i(\boldsymbol{\psi}) = \mathbb{P}(Y_i = 1 \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) &= \mathbb{P}(X_i \le d_i \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) \\ &= \mathbb{P}(X_i \le d_i \mid \boldsymbol{F} = -\boldsymbol{\psi}) \\ &= \mathbb{P}(\sqrt{1-\beta_i}\varepsilon_i \le d_i + \sqrt{\beta_i}\mathbf{a}_i'\boldsymbol{\psi}) \\ &= \Phi\left(\frac{d_i + \sqrt{\beta_i}\mathbf{a}_i'\boldsymbol{\psi}}{\sqrt{1-\beta_i}}\right) \\ &= \Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\beta_i}\mathbf{a}_i'\boldsymbol{\psi}}{\sqrt{1-\beta_i}}\right). \end{aligned}$$

- $p_i(\boldsymbol{\Psi})$ has a probit-normal distribution. The parameters μ_i and σ_i are

$$\mu_i = \Phi^{-1}(p_i)/\sqrt{1-\beta_i} \quad \text{and} \quad \sigma_i^2 = \beta_i/(1-\beta_i)\,.$$

Special cases

1) One-Factor Model

$$p_i(\psi) = \Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\beta_i}\psi}{\sqrt{1-\beta_i}}\right),$$

where Ψ is a standard normally distributed factor.

2) One Factor Model with Equicorrelation Structure

$$p_i(\psi) = \Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\rho}\psi}{\sqrt{1-\rho}}\right), \tag{120}$$

where ρ is the asset correlation between any two critical variables $X_i \neq X_j$.

3) Fully exchangeable model. We set $p_1 = \cdots = p_m$ in equicorrelation model.

11.2.5 Poisson mixture models and CreditRisk+

- Since default is typically a rare event, it is possible to approximate Bernoulli indicator rvs for default with Poisson rvs and Bernoulli mixture models with Poisson mixture models.
- By choosing independent gamma distributions for the economic factors $\boldsymbol{\Psi}$, we obtain a tractable model known as CreditRisk+, proposed by Credit Suisse Financial Products in 1997.
- Assume that, given the factors $\boldsymbol{\Psi}$, the default indicators $Y_1, \dots, Y_m$ for a particular time horizon are conditionally independent Bernoulli variables satisfying $\mathbb{P}(Y_i = 1 \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) = p_i(\boldsymbol{\psi})$.
- Moreover assume that the distribution of $\boldsymbol{\Psi}$ is such that the conditional default probabilities $p_i(\boldsymbol{\psi})$ tend to be very small.
- The Y_i variables can be approximated by conditionally independent Poisson variables $\tilde{Y}_i$ satisfying $\tilde{Y}_i \mid \boldsymbol{\Psi} = \boldsymbol{\psi} \sim \text{Poi}(p_i(\boldsymbol{\psi}))$.

- This follows because

$$\mathbb{P}(\tilde{Y}_i = 0 \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) = e^{-p_i(\boldsymbol{\psi})} \approx 1 - p_i(\boldsymbol{\psi}),$$
$$\mathbb{P}(\tilde{Y}_i = 1 \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) = p_i(\boldsymbol{\psi}) e^{-p_i(\boldsymbol{\psi})} \approx p_i(\boldsymbol{\psi}).$$

- The portfolio loss $L = \sum_{i=1}^m e_i \delta_i Y_i$ can be approximated by $\tilde{L} = \sum_{i=1}^m e_i \delta_i \tilde{Y}_i$.
- It is possible for a company to "default more than once" in the approximating Poisson model, albeit with a very low probability.
- In CreditRisk+ the parameter $\lambda_i(\boldsymbol{\Psi})$ of the conditional Poisson distribution for firm i is assumed to take the form

$$\lambda_i(\boldsymbol{\Psi}) = k_i \boldsymbol{w}_i' \boldsymbol{\Psi} \tag{121}$$

for $k_i > 0$, non-negative weights $\boldsymbol{w}_i = (w_{i1}, \ldots, w_{ip})'$ satisfying $\sum_j w_{ij} = 1$, and p independent $\mathrm{Ga}(\alpha_j, \beta_j)$-distributed factors $\Psi_1, \ldots, \Psi_p$.
- The parameters are set to be $\alpha_j = \beta_j = \sigma_j^{-2}$ for $\sigma_j > 0$ and $j = 1, \ldots, p$.

- This parametrization of the gamma variables ensures that we have $\mathbb{E}(\Psi_j) = 1$ and $\mathrm{var}(\Psi_j) = \sigma_j^2$.
- It is easy to verify that

$$\mathbb{E}(\tilde{Y}_i) = \mathbb{E}(\mathbb{E}(\tilde{Y}_i \mid \boldsymbol{\Psi})) = \mathbb{E}(\lambda_i(\boldsymbol{\Psi})) = k_i \mathbb{E}(\boldsymbol{w}_i' \boldsymbol{\Psi}) = k_i,$$

  so k_i is the expected number of defaults for obligor i in the time period.
- The assumptions in CreditRisk+ make it possible to compute the distribution of the number of defaults and the aggregate portfolio loss fairly explicitly using techniques for compound distributions and mixture distributions that are well known in actuarial mathematics.
- The exchangeable version of CreditRisk+ with $k_i = k, \forall i$, and a single gamma-distributed factor can be shown to be very close to an exchangeable Bernoulli mixture model with beta mixing distribution.

Distribution of the number of defaults

In CreditRisk+ we have that given $\mathbf{\Psi} = \boldsymbol{\psi}$, $\tilde{Y}_i \sim \mathrm{Poi}(k_i \boldsymbol{w}_i' \boldsymbol{\psi})$, which implies that the distribution of the number of defaults $\tilde{M} := \sum_{i=1}^m \tilde{Y}_i$ satisfies

$$\tilde{M} \mid \mathbf{\Psi} = \boldsymbol{\psi} \sim \mathrm{Poi}\left(\sum_{i=1}^{m} k_i \boldsymbol{w}_i' \boldsymbol{\psi}\right). \tag{122}$$

- This uses the fact that the sum of independent Poisson variables is also Poisson with a rate parameter given by the sum of the rate parameters
- To compute the unconditional distribution of $\tilde{M}$ we require a well-known result on mixed Poisson distributions.

Proposition 11.5

If the rv N is conditionally Poisson with a gamma-distributed rate parameter $\Lambda \sim \mathrm{Ga}(\alpha, \beta)$, then N has a negative binomial distribution, $N \sim \mathrm{NB}(\alpha, \beta/(\beta+1))$.

In the case when $p = 1$ we may apply this result directly to (122) to deduce that $\tilde{M}$ has a negative binomial distribution. The general result is:

Proposition 11.6

$\tilde{M}$ is distributed as a sum of p independent negative binomial rvs.

This follows by observing that

$$\sum_{i=1}^{m} k_i \boldsymbol{w}_i' \boldsymbol{\Psi} = \sum_{i=1}^{m} k_i \sum_{j=1}^{p} w_{ij} \Psi_j = \sum_{j=1}^{p} \Psi_j \Big(\sum_{i=1}^{m} k_i w_{ij} \Big).$$

Now consider rvs $\tilde{M}_1, \ldots, \tilde{M}_p$ such that $\tilde{M}_j$ is conditionally Poisson with mean $(\sum_{i=1}^{m} k_i w_{ij})\psi_j$ conditional on $\Psi_j = \psi_j$. The independence of the components $\Psi_1, \ldots, \Psi_p$ implies that the $\tilde{M}_j$ are independent, and by construction we have $\tilde{M} \stackrel{\mathrm{d}}{=} \sum_{j=1}^{p} \tilde{M}_j$. Moreover, the rvs $(\sum_{i=1}^{m} k_i w_{ij})\Psi_j$ are gamma distributed, so that each of the $\tilde{M}_j$ has a negative binomial distribution by Proposition 11.5.

Distribution of the aggregate loss

- To obtain a tractable model, exposures are discretized in CreditRisk+ using the concept of exposure bands.
- The LGD is subsumed in the exposure by multiplying the actual exposure by a typical value for the LGD for an obligor with the same credit rating.
- The losses arising from the individual obligors are of the form $\tilde{L}_i = e_i \tilde{Y}_i$ where the e_i are known (LGD-adjusted) exposures.
- For all i, the exposure e_i is discretized in units of an amount ϵ so that e_i is replaced by a value $\ell_i \epsilon \geq e_i$ where ℓ_i is a positive integer multiplier.
- Exposure bands $b = 1, \ldots, n$ are defined corresponding to the distinct values $\ell^{(1)}, \ldots, \ell^{(n)}$ for the multipliers so that obligors are grouped in exposure bands according to the values of their discretized exposures.
- It is then possible to derive the distribution of the aggregate loss $\tilde{L} = \sum_{i=1}^{m} \ell_i \epsilon \tilde{Y}_i$.

Theorem 11.7

Let $\tilde{L}$ represent the aggregate loss in the general p-factor CreditRisk+ model with exposures discretized into exposure bands as described above. Then the following hold.

i) The Laplace–Stieltjes transform of the df of $\tilde{L}$ is given by

$$\hat{F}_{\tilde{L}}(s) = \prod_{j=1}^{p}\left(1+\sigma_j^2\sum_{i=1}^{m}k_i w_{ij}\left(1-\sum_{b=1}^{n}e^{-s\epsilon\ell^{(b)}}q_{jb}\right)\right)^{-\sigma_j^{-2}}, \tag{123}$$

where $q_{jb} = \sum_{i\in s_b} k_i w_{ij} / \sum_{i=1}^{m} k_i w_{ij}$ for $b = 1, \dots, n$.

ii) The distribution of $\tilde{L}$ has the structure $\tilde{L} \stackrel{d}{=} \sum_{j=1}^{p} Z_j$ where the Z_j are independent variables that follow a compound negative binomial distribution. More precisely, it holds that $Z_j \sim \mathrm{CNB}(\sigma_j^{-2}, \theta_j, G_{X_j})$ with $\theta_j = (1+\sigma_j^2\sum_{i=1}^{m}k_i w_{ij})^{-1}$ and G_{X_j} the df of a multinomial random variable X_j taking the value $\epsilon\ell^{(b)}$ with probability q_{jb}.

11.3 Asymptotics for large portfolios

11.3.1 Exchangeable one-factor models

- We begin the study of asymptotics with the special case of an exchangeable Bernoulli mixture model.
- We consider an infinite sequence of obligors indexed by $i \in \mathbb{N}$ with identical exposures $e_i = e$ and LGD equal to 100%.
- We assume that, given a mixing variable $Q \in [0, 1]$ the default indicators Y_i are independent Bernoulli random variables with conditional default probability $\mathbb{P}(Y_i = 1 \mid Q = q) = q$.
- We are interested in the asymptotic behaviour of the relative loss (the loss expressed as a proportion of total exposure).
- Writing $L^{(m)} = \sum_{i=1}^{m} eY_i$ for the total loss of the first m companies,

the corresponding relative loss is given by

$$\frac{L^{(m)}}{me} = \frac{1}{m} \sum_{i=1}^{m} Y_i \,.$$

- Conditioning on $Q = q$ the Y_i are independent with mean q and the strong law of large numbers (SLLN) implies that, given $Q = q$,

$$\lim_{m \to \infty} \frac{L^{(m)}}{me} = q$$

almost surely.

- This shows that, for large m, the behaviour of the relative loss is essentially governed by the mixing distribution $G(q)$ of Q. In particular, it can be shown that, for G strictly increasing,

$$\lim_{m \to \infty} \mathrm{var}_\alpha \left(\frac{L^{(m)}}{me} \right) = q_\alpha(Q) \,.$$

Example: Vasicek distribution

- Consider the exchangeable model that is equivalent to a Gaussian threshold model with default probability π and default correaltion ρ.
- In this model Q has a probit-normal distribution given by

$$Q = \Phi\left(\frac{\Phi^{-1}(\pi) + \sqrt{\rho}\Psi}{\sqrt{1-\rho}}\right).$$

- The ideas of using this distribution as an approximation to the large portfolio loss is attributed to *Vasicek*.
- Similar asymptotic ideas can be applied to more complicated models. We find in general that the distribution of the loss is driven by the distribution of the *systematic factor* and idiosyncratic risks become neglibible.
- The asymptotic analysis is also described as analysing *infinitely fine-grained portfolios*.

The large portfolio results can be used to give approximations when m is large. Consider again the simple example.

- For tail probabilities we have

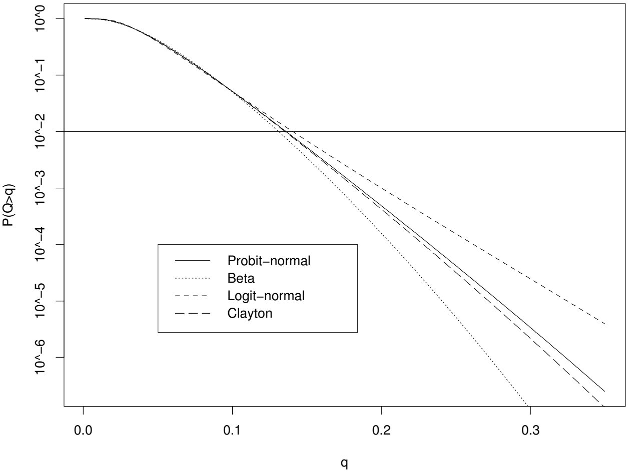
$$\mathbb{P}\left(L^{(m)} > l\right) \approx \mathbb{P}\left(Q > \frac{l}{me}\right) = \Phi\left(\frac{\Phi^{-1}(\pi) - \sqrt{1-\rho}\,\Phi^{-1}(l/(me))}{\sqrt{\rho}}\right).$$

- For value-at-risk we have

$$\mathrm{VaR}_\alpha(L^{(m)}) \approx me \cdot q_\alpha(Q) = me \cdot \Phi\left(\frac{\Phi^{-1}(\pi) + \sqrt{\rho}\,\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right).$$

An important insight is that the tail of distribution of Q determines the tail of portfolio loss distribution.

Comparison of 4 models with same values for π and π_2.



Horizontal line at 0.01 shows models diverge at 99th percentile of Q.

11.3.2 General results

- Let $(e_i)_{i \in \mathrm{N}}$ be an infinite sequence of positive deterministic exposures, $(Y_i)_{i \in \mathrm{N}}$ be the corresponding sequence of default indicators and $(\delta_i)_{i \in \mathrm{N}}$ a sequence of random variables with values in $(0, 1]$ representing percentage losses given that default occurs.
- In this setting the loss for a portfolio of size m is given by $L^{(m)} = \sum_{i=1}^{m} L_i$ where $L_i = e_i \delta_i Y_i$ are the individual losses. We now make some technical assumptions for our model.
- We introduce the notation $a_m = \sum_{i=1}^{m} e_i$ for the aggregate exposure to the first m obligors.
- We now make some technical assumptions for our model.

Assumptions

- A1. There is a p-dimensional random vector $\mathbf{\Psi}$ such that, conditional on $\mathbf{\Psi}$, the $(L_i)_{i\in\mathbb{N}}$ form a sequence of independent random variables.
  We extend conditional independence assumption to losses.
- A2. There is a function $\bar{\ell} : \mathbb{R}^p \to [0,1]$ such that
  $$\lim_{m\to\infty} \frac{1}{a_m}\mathbb{E}(L^{(m)} \mid \mathbf{\Psi} = \boldsymbol{\psi}) = \bar{\ell}(\boldsymbol{\psi})$$
  for all $\boldsymbol{\psi} \in \mathbb{R}^p$. We call $\bar{\ell}(\boldsymbol{\psi})$ the asymptotic relative loss function.
  We preserve composition of portfolio as it grows.
- A3. The sequence of exposures satisfies
  $$\lim_{m\to\infty} a_m = \infty \quad \text{and} \quad \sum_{i=1}^{\infty} (e_i/a_i)^2 < \infty.$$
  The portfolio may not be dominated by a few large exposures.

First result

- We can show that in large portfolios the portfolio loss is essentially determined by the asymptotic relative loss function $\bar{\ell}$ and the realisation of the factor random vector $\boldsymbol{\Psi}$.
- Consider a sequence $L^{(m)} = \sum_{i=1}^{m} L_i$ satisfying the assumptions.
- Denote by $\mathbb{P}(\cdot \mid \boldsymbol{\Psi} = \boldsymbol{\psi})$ the conditional distribution of the sequence $(L_i)_{i \in \mathrm{N}}$ given $\boldsymbol{\Psi} = \boldsymbol{\psi}$. Then

$$\lim_{m \to \infty} \frac{1}{a_m} L^{(m)} = \bar{\ell}(\boldsymbol{\psi}), \quad \mathbb{P}(\cdot \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) \text{ a.s.}$$

- The proof uses a version of the law of large numbers for non-identically-distributed random variable by Petrov (1995).

Second result

- For one-factor Bernoulli mixture models we can obtain a stronger result which links the quantiles of $L^{(m)}$ to quantiles of the mixing distribution.

- Consider a sequence $L^{(m)} = \sum_{i=1}^{m} L_i$ satisfying the assumptions with a one-dimensional mixing variable Ψ with distribution function $G(\psi)$.
- Technical: assume that the conditional asymptotic loss function $\bar{\ell}(\psi)$ is *strictly increasing* and continuous and that G is strictly increasing at $q_\alpha(\Psi)$, i.e. that $G(q_\alpha(\Psi) + \delta) > \alpha$ for every $\delta > 0$.
- Then
$$\lim_{m\to\infty} \mathrm{var}_\alpha\left(\frac{1}{a_m} L^{(m)}\right) = \bar{\ell}(q_\alpha(\Psi)). \tag{124}$$
- The proof is based on the following simple intuition. Since $L^{(m)}/a_m$ converges to $\bar{\ell}(\Psi)$ and since $\bar{\ell}$ is strictly increasing by assumption we have for large m
$$q_\alpha\left(\frac{L^{(m)}}{a_m}\right) \approx q_\alpha(\bar{\ell}(\Psi)) = \bar{\ell}(q_\alpha(\Psi))\,.$$

Application to exchangeable groups

- Consider the one-factor Bernoulli mixture model for k exchangeable groups defined by (119).
- Denote by $r(i)$ the group of obligor i and assume that, within each group r, the exposures, LGDs and conditional default probabilities are identical and given by e_r, δ_r and $p_r(\psi)$ respectively.
- Suppose that we allow the portfolio to grow and write $m_r^{(m)}$ for the number of obligors in group r when the portfolio size is m.
- The relative exposure to group r is given by

$$\lambda_r^{(m)} = e_r m_r^{(m)} / \sum_{r=1}^{k} e_r m_r^{(m)}$$

  and we assume that $\lambda_r^{(m)} \to \lambda_r$ as $m \to \infty$.

- The asymptotic relative loss function is

$$\bar{\ell}(\psi) = \lim_{m\to\infty} \sum_{i=1}^{m} \frac{e_{r(i)}}{\sum_{i=1}^{m} e_{r(i)}} \delta_{r(i)} p_{r(i)}(\psi)$$
$$= \lim_{m\to\infty} \sum_{r=1}^{k} \frac{e_r m_r^{(m)}}{\sum_{r=1}^{k} e_r m_r^{(m)}} \delta_r h(\mu_r + \sigma\psi) = \sum_{r=1}^{k} \lambda_r \delta_r h(\mu_r + \sigma\psi)\,.$$

- Since Ψ is assumed to have a standard normal distribution, (124) implies that

$$\lim_{m\to\infty} q_\alpha \left(\frac{L^{(m)}}{\sum_{i=1}^{m} e_i} \right) = \sum_{r=1}^{k} \lambda_r \delta_r h(\mu_r + \sigma\Phi^{-1}(\alpha)). \qquad (125)$$

- For large m, since $\lambda_r \sum_{1=1}^{m} e_i \approx m_r^{(m)} e_r$, we get that

$$\mathrm{var}_\alpha(L^{(m)}) \approx \sum_{r=1}^{k} m_r^{(m)} e_r \delta_r h(\mu_r + \sigma_r \Phi^{-1}(\alpha)). \qquad (126)$$

- This formula is the basis of the Basel IRB formula.

11.3.3 The Basel IRB formula

- Under the Basel framework a bank is required to hold 8% of the so-called *risk-weighted assets* (RWA) of its credit portfolio as risk capital.
- The RWA of a portfolio is given by the sum of the RWA of the individual risks in the portfolio, i.e. $\text{RWA}^{\text{portfolio}} = \sum_{i=1}^{m} \text{RWA}_i$.
- The quantity RWA_i reflects the exposure size and riskiness of obligor i and takes the form $\text{RWA}_i = w_i e_i$, where w_i is a risk weight and e_i denotes exposure size.
- Banks may choose between two options for determining the risk weight w_i, which must then be implemented for the entire portfolio. Under the simpler *standardized approach*, the risk weight w_i is determined by the type (sovereign, bank or corporate) and the credit rating of counterparty i.
- For instance, $w_i = 50\%$ for a corporation with a Moody's rating in the range of A$+$ to A$-$.

- Under the *internal-ratings-based* (IRB) approach, the risk weight is determined by a formula where inputs may be determined by the bank.
- The IRB risk weights takes the form

$$w_i = (0.08)^{-1} c\delta_i \Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\beta_i}\Phi^{-1}(0.999)}{\sqrt{1-\beta_i}}\right).$$

- Here c is a technical adjustment factor, p_i represents the default probability, and δ_i is the percentage loss given default of obligor i.
- $\beta_i \in (0.12, 0.24)$ measures the systematic risk of obligor i.
- Estimates for p_i and (under the so-called advanced IRB approach) for δ_i and e_i are provided by the individual bank.
- The adjustment factor c and, most importantly, the value of β_i are determined by fixed rules within the Basel II Accord.
- The risk capital to be held for counterparty i is thus given by

$$\mathrm{RC}_i = 0.08\mathrm{RWA}_i = c\delta_i e_i \Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\beta_i}\Phi^{-1}(0.999)}{\sqrt{1-\beta_i}}\right). \quad (127)$$

11.4 Monte Carlo methods

11.4.1 Basics of importance sampling

In a generic Monte Carlo problem we have an rv X with density f and we wish to compute an expected value of the form

$$\theta = \mathbb{E}(h(X)) = \int_{-\infty}^{\infty} h(x)f(x)dx, \tag{128}$$

for some known function h. For an event probability $h(x) = I_{\{x \in A\}}$ for some set $A \subset \mathbb{R}$; for expected shortfall computation $h(x) = xI_{\{x \geq c\}}$ where $c = \mathrm{VaR}_\alpha$.

Where analytical evaluation of θ is difficult we can use an MC approach:

1) Simulate $X_1, \ldots, X_n$ independently from density f.

2) Compute the standard MC estimate $\hat{\theta}_n^{\mathsf{MC}} = \frac{1}{n}\sum_{i=1}^n h(X_i)$.

- The MC estimator converges to θ by the strong law of large numbers, (SLLN) but the speed of convergence may not be particularly fast, particularly when we are dealing with rare event simulation.
- Suppose we want to estimate $\mathrm{ES}_{0.99}(L)$ in credit risk context. Only 1% of our standard Monte Carlo draws will lead to a portfolio loss higher than $\mathrm{VaR}_{0.99}(L)$. The standard MC estimator of, which consists of averaging the simulated values of L over all draws leading to a simulated portfolio loss $L \geq \mathrm{VaR}_{\alpha}(L)$, will be unstable and subject to high variability, unless the number of simulations is very large.
- The technique of *importance sampling* is a way of reducing this variability and is well suited to problems of the kind we consider.
- Importance sampling is based on an alternative representation of θ in (128).

Importance sampling algorithm

- We consider an *importance sampling density* g (whose support should contain that of f) and define the *likelihood ratio* $r(x)$ by $r(x) := f(x)/g(x)$ whenever $g(x) > 0$ and $r(x) = 0$ otherwise.
- The integral may be written as

$$\theta = \int_{-\infty}^{\infty} h(x)r(x)g(x)dx = \mathbb{E}_g(h(X)r(X)), \tag{129}$$

where $\mathbb{E}_g$ denotes expectation with respect to the density g.

We can approximate the integral with the following algorithm;

1) Simulate $X_1, \ldots, X_n$ independently from density g.
2) Compute the IS estimate $\hat{\theta}_n^{\mathrm{IS}} = \frac{1}{n} \sum_{i=1}^{n} h(X_i)\, r(X_i)$.

Reducing the variance

- The art of importance sampling is in choosing g such that for fixed n the variance of the IS estimator is considerably smaller than that of the standard Monte Carlo estimator.

$$\begin{aligned}\operatorname{var}_g\left(\hat{\theta}_n^{\mathsf{IS}}\right) &= \frac{\mathbb{E}_g\left(h(X)^2 r(X)^2\right)-\theta^2}{n}=\frac{\mathbb{E}\left(h(X)^2 r(X)\right)-\theta^2}{n},\\ \operatorname{var}\left(\hat{\theta}_n^{\mathsf{MC}}\right) &= \frac{\mathbb{E}\left(h(X)^2\right)-\theta^2}{n}.\end{aligned}$$

- The aim is to make $\mathbb{E}(h(X)^2 r(X))$ small compared to $\mathbb{E}(h(X)^2)$.
- Consider the case of estimating a tail probability where $h(x)=I_{\{x\geq c\}}$ for c significantly larger than the mean of X. We try to choose g so that the likelihood ratio $r(x)=f(x)/g(x)$ is small for $x\geq c$; in other words we make the event $\{X\geq c\}$ more likely under the IS density g than it is under the original density f.

- A technique for construcing importance sampling densities is known as *exponential tilting*.
- For $t \in \mathbb{R}$ write $M_X(t) = \mathbb{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)dx$ for the moment generating function of X, which we assume is finite for $t \in \mathbb{R}$.
- We can define a density $g_t(x) = e^{tx} f(x)/M_X(t)$ which can be used for importance sampling when X is light tailed.
- The likelihood ratio is $r_t(x) = f(x)/g_t(x) = M_X(t)e^{-tx}$.
- Define μ_t to be the mean of X with respect to the density g_t i.e.

$$\mu_t := \mathbb{E}_{g_t}(X) = \mathbb{E}\big(X \, \exp(tX)/M_X(t)\big).$$

- How can we choose t optimally for a particular importance sampling problem?
- In the case of tail probability estimation theory suggests we should choose t as the solution of $\mu_t = c$.

Example 11.8 (Exponential tilting for normal distribution)

- We illustrate the concept of exponential tilting in the simple case of a standard normal random variable. Suppose $X \sim \mathrm{N}(0,1)$ with density $\phi(x)$.
- Using exponential tilting we obtain the new density

$$g_t(x) = \exp(tx)\phi(x)/M_X(t).$$

The moment generating function of X is known to be $M_X(t) = \exp(t^2/2)$.

- Hence

$$g_t(x) = \frac{1}{\sqrt{2\pi}} \exp\left(tx - \frac{1}{2}(t^2 + x^2)\right) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{(x-t)^2}{2}\right),$$

so that under the tilted distribution, $X \sim \mathrm{N}(t,1)$.

- Exponential tilting is a convenient way of shifting the mean of X.

An abstract view of importance sampling

To handle the more complex application to portfolio credit risk we consider importance sampling from a slightly more general viewpoint.

- Given densities f and g we define probability measures $\mathbb{P}$ and $\mathbb{Q}$ by
$$\mathbb{P}(A) = \int_A f(x)dx \quad \text{and} \quad \mathbb{Q}(A) = \int_A g(x)dx\,,\ A \subset \mathbb{R}\,.$$
- With this notation (129) becomes
$$\theta = \mathbb{E}_{\mathbb{P}}(h(X)) = \mathbb{E}_{\mathbb{Q}}(h(X)r(X)),$$
so that $r(X)$ equals $d\mathbb{P}/d\mathbb{Q}$, the (measure-theoretic) density of $\mathbb{P}$ with respect to $\mathbb{Q}$.
- Using this more abstract view, exponential tilting can be applied in more general situations.

- Given a rv X on $(\Omega, \mathcal{F}, P)$ such that $M_X(t) = \mathbb{E}_{\mathbb{P}}(\exp(tX)) < \infty$, define the measure $\mathbb{Q}_t$ on $(\Omega, \mathcal{F})$ by

$$\frac{d\mathbb{Q}_t}{d\mathbb{P}} = \frac{\exp(tX)}{M_X(t)}$$

  and note that $(d\mathbb{Q}_t/d\mathbb{P})^{-1} = r_t(X)$.

- Event probabilities are now calculated according to

$$\mathbb{Q}_t(A) = \mathbb{E}_{\mathbb{Q}}(I_A) = \mathbb{E}_{\mathbb{P}}(r_t(X)^{-1} I_A) = \mathbb{E}_{\mathbb{P}}\left(\frac{\exp(tX)}{M_X(t)} I_A\right),$$

- The IS algorithm remains essentially unchanged: simulate independent realizations X_i under the measure $\mathbb{Q}_t$ and set $\hat{\theta}^{\mathsf{IS}} = \frac{1}{n}\sum_{i=1}^{n} X_i r_t(X_i)$ as before.

11.4.2 Application to Bernoulli-mixture models

- Consider a portfolio loss of the form $L = \sum_{i=1}^{m} e_i Y_i$, where the e_i are deterministic, positive exposures and the Y_i are default indicators with default probabilities p_i. $\mathbf{Y}$ follows a Bernoulli mixture model with factor vector $\mathbf{\Psi}$ and conditional default probabilities $p_i(\mathbf{\Psi})$.
- We study the problem of estimating $\theta = \mathbb{P}(L \geq c)$ for c substantially larger than $\mathbb{E}(L)$ using importance sampling.
- We consider first the situation where the default indicators $Y_1, \ldots, Y_m$ are independent and discuss subsequently the extension to the case of conditionally independent default indicators.
- Here we use the more general IS approach and set $\Omega = \{0,1\}^m$, the state space of $\mathbf{Y}$. The probability measure $\mathbb{P}$ is given by

$$\mathbb{P}(Y_1 = y_1, \ldots, Y_m = y_m) = \prod_{i=1}^{m} p_i^{y_i}(1 - p_i)^{1-y_i}, \quad (y_1, \ldots, y_m) \in \Omega.$$

- The moment generating function of L is
$$M_L(t) = \mathbb{E}\left(\exp\left(t\sum_{i=1}^m e_i Y_i\right)\right) = \prod_{i=1}^m \mathbb{E}\left(e^{te_iY_i}\right) = \prod_{i=1}^m \left(e^{te_i}p_i + 1 - p_i\right).$$

- Under the new measure $\mathbb{Q}_t$ we have
$$\begin{aligned}\mathbb{Q}_t(Y_1 = y_1, \dots, Y_m = y_m) &= \mathbb{E}_{\mathbb{P}}\left(\frac{e^{tL}}{M_L(t)} I_{\{Y_1=y_1,\dots,Y_m=y_m\}}\right)\\ &= \frac{e^{\left(t\sum_{i=1}^m e_i y_i\right)}}{M_L(t)}\mathbb{P}(Y_1 = y_1, \dots, Y_m = y_m).\end{aligned}$$

- We obtain
$$= \prod_{i=1}^m \frac{\exp(te_iy_i)}{\exp(te_i)p_i + 1 - p_i}\, p_i^{y_i}(1-p_i)^{1-y_i}.$$

- Define $q_{t,i} := \exp(te_i)p_i/(\exp(te_i)p_i + 1 - p_i)$.

- It follows that
$$\mathbb{Q}_t(Y_1 = y_1, \dots, Y_m = y_m) = \prod_{i=1}^m q_{t,i}^{y_i}(1-q_{t,i})^{1-y_i}$$

- We see that the default indicators remain independent but with new default probability $q_{t,i}$.
- The optimal value of t is chosen such that $\mathbb{E}_{\mathbb{Q}_t}(L) = c$, leading to the equation $\sum_{i=1}^m e_i q_{t,i} = c$.

Conditionally independent defaults

- The first step is obvious: given a realization $\boldsymbol{\psi}$ of the economic factors, the conditional exceedance probability $\theta(\boldsymbol{\psi}) := \mathbb{P}(L \geq c \mid \boldsymbol{\Psi} = \boldsymbol{\psi})$ is estimated using the approach for independent default indicators.
- This gives an estimate $\hat{\theta}_{n_1}^{\mathsf{IS},1}(\boldsymbol{\psi})$ where n_1 is the number of random draws of $(Y_1, \ldots, Y_m)$.
- Our aim is to estimate

$$\theta = \mathbb{P}(L \geq c) = \mathbb{E}\left(\mathbb{P}(L \geq c \mid \boldsymbol{\Psi}\right) = \mathbb{E}\left(\theta(\boldsymbol{\Psi})\right).$$

- In a naive approach we could generate n realizations of $\boldsymbol{\Psi}$ and estimate θ by calculating the average $\frac{1}{n}\sum_{i=1}^n \hat{\theta}_{n_1}^{\mathsf{IS},1}(\boldsymbol{\Psi}_i)$.

- However a dramatic improvement can be obtained by also applying importance sampling to the distribution of $\mathbf{\Psi}$.
- We will illustrate this idea with a one-factor example.
- Consider the one-factor Gaussian threhold model with conditional default probabilities $p_i(\Psi)$ where $\Psi \sim \mathrm{N}(0,1)$.
- Instead of generating $\Psi_1, \ldots, \Psi_n$ from a standard normal $\mathrm{N}(0,1)$ distribution we should use exponential tilting to generate them from a $\mathrm{N}(\mu,1)$ distribution for some sensibly chosen value of μ.
- Note that we will not discuss the optimal choice of μ; see Glasserman and Li (2005) for more details.

We obtain the algorithm:

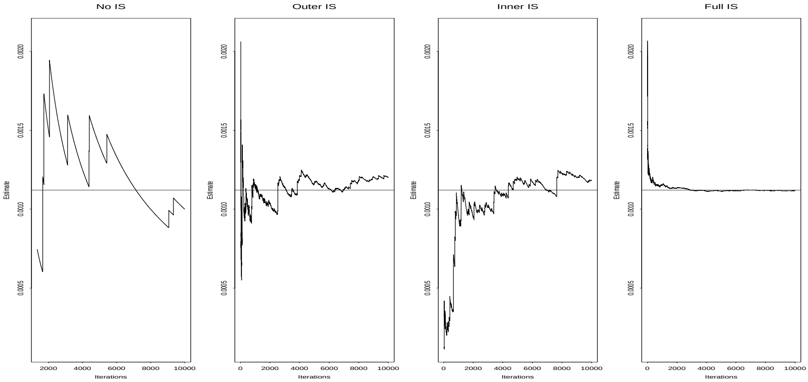
1) Generate $\Psi_1, \ldots, \Psi_n \sim \mathrm{N}(\mu,1)$ independently.
2) For each Ψ_i calculate $\hat{\theta}_{n_1}^{\mathsf{IS},1}(\Psi_i)$ by importance sampling.

3) Determine the full IS estimator: $\hat{\theta}_n^{\mathsf{IS}} = \frac{1}{n}\sum_{i=1}^{n} r_\mu(\Psi_i)\hat{\theta}_n^{\mathsf{IS},1}(\Psi_i)$ where $r_\mu(\psi) = \exp(-\mu\psi + \frac{1}{2}\mu^2)$.

Example 11.9 (Exchangeable portfolio)

- We consider an exchangeable portfolio of 100 firms with identical unit exposures, default probabilities 0.05 and asset correlations (i.e. values of ρ) 0.05.
- Aim is to calculate the tail probability $\mathbb{P}(L \geq 20)$ by IS.
- In such a simple model it can in fact be calculated analytically to be 0.00112.
- We compare the following MC/IS methods:
  1) Naive Monte Carlo ($n = 10000$) (No IS)
  2) IS for factor distribution ($n = 10000$) (outer IS)
  3) Naive Monte Carlo for factor ($n = 10000$) and IS for conditional default distribution ($n1 = 50$) (inner IS)

4) IS for factor distribution ($n = 10000$) and conditional default distribution ($n1 = 50$) (Full IS)



11.5 Statistical inference for portfolio credit models

11.5.1 Industry factor models

- Recall that portfolio models in industry often take the form of a Gaussian threshold model $(\boldsymbol{X}, \boldsymbol{d})$ with $\boldsymbol{X} \sim \mathrm{N}_m(\mathbf{0}, P)$, where the random vector $\boldsymbol{X}$ contains the critical variables, the deterministic vector $\boldsymbol{d}$ contains the critical default thresholds and P is the so-called asset correlation matrix, which is estimated with the help of a factor model for $\boldsymbol{X}$.
- Industry models generally separate the calibation of the vector $\boldsymbol{d}$ (or the threshold matrix D in a multi-state model) and the calibration of the factor model for $\boldsymbol{X}$.
- In a default-only model the threshold d_i is usually set at $d_i = \Phi^{-1}(p_i)$ where p_i is an estimate of the default probability for obligor i for the time period in question (generally one year).

- The default probability may be estimated in different ways: for larger corporates it may be estimated using credit ratings or using a firm-value approach, such as the Moody's public-firm EDF model; for retail obligors it may be estimated on the basis of credit scores.
- Recall that the factor model for $\boldsymbol{X}$ takes the form

$$X_i = \sqrt{\beta_i}\tilde{F}_i + \sqrt{1-\beta_i}\varepsilon_i, \quad i = 1, \dots, m, \tag{130}$$

  where $\tilde{F}_i$ and $\varepsilon_1, \dots, \varepsilon_m$ are independent standard normal variables, and where $0 \leq \beta_i \leq 1$ for all i.
- The systematic variables $\tilde{F}_i$ are assumed to be of the form $\tilde{F}_i = \boldsymbol{a}_i'\boldsymbol{F}$ where $\boldsymbol{F}$ is a vector of common factors satisfying $\boldsymbol{F} \sim \mathrm{N}_p(\mathbf{0}, \Omega)$ with $p < m$, and where Ω is a correlation matrix.
- The factors typically represent country and industry effects.
- The assumption that $\mathrm{var}(\tilde{F}_i) = 1$ implies that $\boldsymbol{a}_i'\Omega\boldsymbol{a}_i = 1$ for all i.

- Different industry models use different data for $\boldsymbol{X}$ to calibrate the factor model (130).
- The Moody's Analytics Global Correlation or GCorr model has sub-models for many different kinds of obligor including public corporate firms, private firms, small and medium enterprises (SMEs), retail customers and sovereigns. Huang et al. (2012)
- The sub-model for public firms (GCorr Corporate) is calibrated using data on weekly asset value returns, where asset values are determined as part of the public-firm EDF methodology.
- In the CreditMetrics framework weekly equity returns are viewed as a proxy for asset returns and used to estimate the factor model.
- We sketch a generic procedure for estimating a factor model for corporates where the factors have country and industry-sector interpretations.

Estimating a credit risk factor model

- We assume that we have a high-dimensional multivariate time series $(\boldsymbol{X}_t)_{1\leq t\leq n}$ of asset returns (or other proxy data for changing credit quality) over a period of time in which stationarity can be assumed.
- We also assume that each component time series has been scaled to have mean zero and variance one.

1) We first fix the structure of the factor vector $\boldsymbol{F}$ so that, for example, the first block of components might represent country factors and the second block of components might represent industry factors. We then assign vectors of factor weights $\boldsymbol{a}_i$ to each obligor based on our knowledge of the companies. The elements of $\boldsymbol{a}_i$ may simply consist of ones and zeros if the company can be clearly identified with a single country and industry, but may also consist of weights if the company has significant activity in more than one country or sector.

2) We then use cross-sectional estimation techniques to estimate the factor values $\boldsymbol{F}_t$ at each time point t. Effectively the factor estimates $\hat{\boldsymbol{F}}_t$ are constructed as weighted sums of the $X_{t,i}$ data for obligors i that are exposed to each factor. One way of achieving this is to construct a matrix A with rows $\boldsymbol{a}_i$ and then to estimate a fundamental factor model of the form $\boldsymbol{X}_t = A\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t$ at each time point t.

- We have a regression model

$$\boldsymbol{X}_t = A\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t, \tag{131}$$

  where $\boldsymbol{X}_t \in \mathbb{R}^m$ are the return data, $A \in \mathbb{R}^{m \times p}$ is a known matrix of factor loadings, $\boldsymbol{F}_t \in \mathbb{R}^p$ are the factors to be estimated and $\boldsymbol{\varepsilon}_t$ are errors with diagonal covariance matrix Υ.

- Note that the components of the error vector $\boldsymbol{\varepsilon}_t$ can not generally be assumed to have equal variance, so that (131) is a regression problem with so-called heteroskedastic errors.

- Unbiased estimators of the factors $\boldsymbol{F}_t$ may be obtained by forming the ordinary least squares (OLS) estimates
$$\hat{\boldsymbol{F}}_t^{\mathsf{OLS}} = (A'A)^{-1}A'\boldsymbol{X}_t\,.$$
- Since the errors are heteroskedastic, slightly more efficient estimators can be obtained by using the method of generalized least squares (GLS).

3) The raw factor estimates form a multivariate time series of dimension p. We standardize each component series to have mean zero and variance one to obtain $(\hat{\boldsymbol{F}}_t)_{1\le t\le n}$ and calculate the sample covariance matrix of the standardized factor estimates, which serves as our estimate of Ω.
4) We then scale the vectors of factor weights $\boldsymbol{a}_i$ so that the conditions $\boldsymbol{a}_i'\hat{\Omega}\boldsymbol{a}_i = 1$ are met for each obligor.
5) Time series of estimated systematic variables for each obligor are then constructed by calculating $\hat{\tilde{F}}_{t,i} = \boldsymbol{a}_i'\hat{\boldsymbol{F}}_t$ for $t = 1,\dots,n$.

6) Finally we estimate the β_i parameters by performing a time series regression of $X_{t,i}$ on $\hat{\hat{F}}_{t,i}$ for each obligor.

Note that the accurate estimation of the β_i in the last step is particularly important (as it effects tail behaviour). The estimate of β_i is the so-called R-squared of the time series regression model in Step 6 and will be largest for the firms whose credit-quality changes are best explained by systematic factors.

11.5.2 Exchangeable Bernoulli-mixture models

- We discuss the estimation of default probabilities and default correlations for homogeneous groups, e.g. groups with the same credit rating.
- Suppose that we observe historical default numbers over n periods of time for a homogeneous group; typically these might be yearly data.
- For $t = 1, \ldots, n$, let m_t denote the number of observed companies at the start of period t.

- Let M_t denote the number that defaulted during the period; the former will be treated as fixed at start of each period and the latter as an rv.
- Suppose further that within a time period these defaults are generated by an exchangeable Bernoulli mixture model.
- In other words, assume that, given some mixing variable Q_t taking values in $(0, 1)$ and the cohort size m_t, the number of defaults M_t is conditionally binomially distributed and satisfies $M_t \mid Q_t = q \sim B(m_t, q)$.
- Further assume that the mixing variables $Q_1, \ldots, Q_n$ are identically distributed.
- We consider using a simple method of moments to estimate the fundamental parameters of the mixing distribution $\pi = \pi_1$, π_2 and ρ_Y (default correlation).
- For $1 \leq t \leq n$, let $Y_{t,1}, \ldots, Y_{t,m_t}$ be default indicators for the m_t

companies in the cohort. Suppose we define the rv

$$Z_{t,k} := \sum_{\{i_1,\ldots,i_k\}\subset\{1,\ldots,m_t\}} Y_{t,i_1} \cdots Y_{t,i_k} \tag{132}$$

which represents the number of possible subgroups of k obligors among the defaulting obligors in period t.

- By taking expectations in (132) we get

$$\mathbb{E}(Z_{t,k}) = \binom{m_t}{k}\pi_k.$$

- Note that we can write

$$Z_{t,k} = \binom{M_t}{k} I_{\{M_t \geq k\}},$$

the number of subgroups of size k among the defaulting obligors.

- We can estimate the unknown theoretical moment π_k by taking a natural empirical average (133) constructed from the n years of data:

$$\hat{\pi}_k = \frac{1}{n}\sum_{t=1}^{n}\frac{Z_{t,k}}{\binom{m_t}{k}} = \frac{1}{n}\sum_{t=1}^{n}\frac{M_t(M_t-1)\cdots(M_t-k+1)}{m_t(m_t-1)\cdots(m_t-k+1)}. \tag{133}$$

- For $k=1$ we get the standard estimator of default probability

$$\hat{\pi} = \frac{1}{n}\sum_{t=1}^{n}\frac{M_t}{m_t},$$

and ρ_Y can obviously be estimated by taking

$$\hat{\rho}_Y = \frac{\hat{\pi}_2 - \hat{\pi}^2}{\hat{\pi} - \hat{\pi}^2}.$$

- The estimator is unbiased for π_k and consistent as $n \to \infty$.
- For more details see Frey and McNeil (2001).