Dependence Modelling, Model Risk and Model Calibration in Models of Portfolio Credit Risk

Rüdiger Frey*
Department of Mathematics
University of Leipzig
Augustusplatz 10/11
D-04109 Leipzig
frey@mathematik.uni-leipzig.de

Alexander J. McNeil*
Department of Mathematics
Federal Institute of Technology
ETH Zentrum
CH-8092 Zurich
mcneil@math.ethz.ch

Abstract

We consider mathematical models for portfolio credit risk. We analyze the mathematical structure and in particular the modelling of dependence between default events in these models and propose extensions of standard industry models. We study the model risk related to the choice of model structure and input parameters. Finally we develop and test several approaches to model calibration in credit risk models.

J.E.L. Subject Classification: G31, G11, C15

Keywords: Risk Management, Credit Risk, Dependence Modelling, Copulas

1 Introduction

A major cause of concern in managing the credit risk in the lending portfolio of a typical financial institution is the occurrence of disproportionately many joint defaults of different counterparties over a fixed time horizon. Joint default events also have an an important impact on the performance of derivative securities, whose payoff is linked to the loss of a whole portfolio of underlying bonds or loans such as collaterized debt obligations (CBOs, CDOs, CLOs) or basket credit derivatives. In fact, the occurrence of disproportionally many joint defaults is what could be termed "extreme credit risk" in these contexts. Clearly, precise mathematical models for the loss in a portfolio of dependent credit risks are needed to adequately measure this risk. Such models are also a prerequisite for the active management of credit portfolios under risk-return considerations. Moreover, given improved availability of data on credit losses, refined versions of current credit risk models might also be used for the determination of regulatory capital for credit risk, much as internal models are nowadays used for capital adequacy purposes in market risk management.

The specification of a model for portfolio credit risk, which is able to capture extreme credit risk, is a major challenge. The model structure has to be flexible enough so that the model is capable of reproducing realistic patterns of joint defaults; moreover, the model parameters have to be estimated properly, which is particularly difficult given the relative scarcity of reliable data on credit losses. These issues are the topic of the present paper. In the first part of the paper we study the modelling of dependence between defaults in models for portfolio credit risk. We analyze the mathematical structure of existing industry models, provide new links between them, discuss potential weaknesses, and develop several extensions and improvements. As a byproduct we obtain new methods for simulating some of the

*We wish to thank Dirk Tasche, Mark Nyfeler, Filip Lindskog, Uwe Schmock, Philipp Schönbucher and Hans Bühlmann for useful discussions. Comments from an anonymous referee were very helpful in sharpening the focus of the paper

models. In this part of the paper we build on recent research on dependence-modelling in risk management in general (see Embrechts, McNeil, and Straumann (2001)); in particular, the copula concept and the notion of tail dependence of risk factors play an important role in our analysis. In the second, more practically oriented part of the paper we study model calibration and model risk (the sensitivity of the distribution of credit losses with respect to parameters and structure of credit models). We find that some popular industry models are subject to a large amount of model risk: a relatively minor change in the structure of the model can have a substantial impact on the tail of the credit loss distribution. Obviously this is a cause of concern for every risk manager using these models. Effective calibration procedures can help to mitigate model risk. We therefore develop new approaches to estimate critical input parameters in two models, which we find particularly useful for practical purposes.

In this paper we focus on models for credit risk management in lending portfolios. These models are closely related to models for the pricing of basket credit derivatives. However, they differ in two important respects: First, in credit risk management one is typically interested in the loss of a portfolio over a fixed time horizon, so that credit portfolio models are static or at least discrete time, whereas models for the pricing of basket credit derivatives are usually set up in continuous time. Second, in credit risk management one is usually interested in the distribution of credit losses under the historical probability measure, whereas the pricing of credit derivatives is usually done using equivalent martingale measures. Hence models for pricing credit derivatives are usually calibrated to market prices of traded credit-sensitive securities (if possible), whereas in the calibration of models used in risk management one relies more on statistical approaches applied to historical default data.

The recent literature contains a number of related papers. Gordy (2000) was the first to analyze the mathematical structure of the most popular credit risk models used in industry, and some of our results on the links between existing industry models build on his stimulating work. A detailed description of popular industry models is given in Crouhy, Galai, and Mark (2000). Related work on pricing basket credit derivatives includes Davis and Lo (2001), Jarrow and Yu (2001), Schönbucher and Schubert (2001), and Giesecke (2001). The common theme of these papers is to construct models which reproduce realistic patterns of joint defaults. The last two papers make explicit use of the copula concept, whereas the papers by Davis and Lo and by Jarrow and Yu propose interesting models for the dynamics of default correlation.

2 Models for loan portfolios

Credit portfolio models currently in use can be divided into two classes. In the statistics literature such as Joe (1997) these are referred to as latent variable models and mixture models. In the latent variable models default occurs if a random variable X (termed latent variable, even if in some models X may be observable) falls below some threshold. Dependence between defaults is caused by dependence between the corresponding latent variables. Popular examples include the firm-value model of Merton (Merton 1974) or the models proposed by the KMV corporation (KMV-Corporation 1997) or the RiskMetrics group (RiskMetrics-Group 1997). In the mixture models the default probability of a company is assumed to depend on a set of economic factors; given these factors, defaults of the individual obligors are conditionally independent. Examples include CreditRisk$^+$, developed by Credit Suisse Financial Products (Credit-Suisse-Financial-Products 1997) and more generally the reduced form models from the credit derivatives literature such as Lando (1998) or Duffie and Singleton (1999). We study latent variable models in Section 3; mixture models and the relation between the two model classes will be considered in Section 4. In the remainder of this section we introduce some general notation.

Consider a portfolio of m obligors. Following the literature on credit risk management we restrict ourselves to static models for most of the analysis; multiperiod models will be considered in Section 5. Fix some time horizon T. For $1 \leq i \leq m$, let the random variable (rv) S_i be a state indicator for obligor i at time T. Assume that S_i takes integer values in the set $\{0, 1, \dots, n\}$ representing for instance rating classes; we interpret the value 0 as default and non-zero values represent states of increasing creditworthiness. At time $t = 0$ obligors are assumed to be in some non-default state. Often we will concentrate on the binary outcomes of default and non-default and ignore the finer categorization of non-defaulted companies. In this case we write Y_i for the default indicator variables; $Y_i = 1 \iff S_i = 0$ and $Y_i = 0 \iff S_i > 0$. The random vector $\mathbf{Y} = (Y_1, \dots, Y_m)'$ is a vector of default indicators for the portfolio and

$$p(\mathbf{y}) = P(Y_1 = y_1, \dots, Y_m = y_m), \quad \mathbf{y} \in \{0, 1\}^m,$$

is its joint probability function; the marginal default probabilities will be denoted by $\overline{p}_i = P(Y_i = 1),\ i = 1, \dots, m$.

We count the number of defaulted obligors at time T with the random variable $M := \sum_{i=1}^m Y_i$. The actual loss if company i defaults – termed loss given default in practice – is modelled by the random quantity $\Delta_i e_i$ where e_i represents the overall exposure to company i and $\Delta_i \in [0, 1]$ represents a random proportion of the exposure which is lost in the default event. We will denote the overall loss by $L := \sum_{i=1}^m e_i \Delta_i Y_i$ and make further assumptions about the e_i's and Δ_i's as and when we need them.

It is possible to set up different credit risk models leading to the same multivariate distribution of $\mathbf{S}$ or $\mathbf{Y}$. Since this distribution is the main object of interest in the analysis of portfolio credit risk, we call two models with state vectors $\mathbf{S}$ and $\widetilde{\mathbf{S}}$ resp. $\mathbf{Y}$ and $\widetilde{\mathbf{Y}}$ *equivalent* if $\mathbf{S} \stackrel{\mathrm{d}}{=} \widetilde{\mathbf{S}}$ resp. $\mathbf{Y} \stackrel{\mathrm{d}}{=} \widetilde{\mathbf{Y}}$, where $\stackrel{\mathrm{d}}{=}$ stands for equality in distribution.

To simplify the analysis we will often assume that the state indicator $\mathbf{S}$ and thus the default indicator $\mathbf{Y}$ is *exchangeable*. This seems the correct way to mathematically formalise the notion of *homogeneous* groups that is used in practice. Recall that a random vector $\mathbf{S}$ is called exchangeable if $(S_1, \dots, S_m) \stackrel{\mathrm{d}}{=} (S_{\Pi(1)}, \dots, S_{\Pi(m)})$ for any permutation $(\Pi(1), \dots, \Pi(m))$ of $(1, \dots, m)$. This implies in particular that for any $k \in \{1, \dots, m-1\}$ all of the $\binom{m}{k}$ possible k-dimensional marginal distributions of $\mathbf{S}$ are identical. In this situation we introduce the following simple notation for default probabilities and joint default probabilities.

$$\begin{aligned} \pi_k &:= P(Y_{i_1} = 1, \dots, Y_{i_k} = 1), \qquad \{i_1, \dots, i_k\} \subset \{1, \dots, m\},\ 1 \leq k \leq m, \\ \pi := \pi_1 &= P(Y_i = 1), \qquad i \in \{1, \dots, m\}. \end{aligned}$$

Thus π_k, the kth order (joint) default probability, is the probability that an arbitrarily selected subgroup of k companies defaults in $[0, T]$. When default indicators are exchangeable we can calculate easily that

$$\begin{aligned} E(Y_i) &= E(Y_i^2) = P(Y_i = 1) = \pi, \quad \forall i, \\ E(Y_i Y_j) &= P(Y_i = 1, Y_j = 1) = \pi_2, \quad i \neq j, \\ \operatorname{cov}(Y_i, Y_j) &= \pi_2 - \pi^2 \text{ and hence } \rho(Y_i, Y_j) = \rho_Y := \frac{\pi_2 - \pi^2}{\pi - \pi^2}, \quad i \neq j. \end{aligned} \tag{1}$$

In particular, the correlation between default indicators is a simple function of the first and second order default probabilities.

3 Latent variables models

3.1 General structure and relation to copulas

Definition 3.1. Let $\mathbf{X} = (X_1, \dots, X_m)'$ be an m-dimensional random vector. For $i \in \{1, \dots, m\}$ let $-\infty = D_{-1}^i < D_0^i < \cdots < D_n^i = \infty$ be a sequence of *cut-off* levels. Set

$$S_i = j \iff X_i \in (D_{j-1}^i, D_j^i] \quad j \in \{0, \dots, n\},\ i \in \{1, \dots, m\}.$$

Then $\left(X_i, (D_j^i)_{-1 \le j \le n}\right)_{1 \le i \le m}$ is a latent variable model for the state vector $\mathbf{S} = (S_1, \dots, S_m)'$.

X_i and D_0^i are often interpreted as the values of assets respectively liabilities for an obligor i at time T; in this interpretation default (corresponding to the event $S_i = 0$) occurs if the value of a company's assets at T is below the value of its liabilities at time T. This modelling of default goes back to Merton (1974) and popular examples incorporating this type of modelling are presented below. We denote by $F_i(x) = P(X_i \le x)$ the marginal distribution functions (df) of $\mathbf{X}$. Obviously, the default probability of company i is given by $\overline{p}_i = F_i(D_0^i)$.

We now give a criterion for equivalence of two latent variable models in terms of the marginal distributions of the state vector $\mathbf{S}$ and the *copula* of $\mathbf{X}$; this result will be very useful in studying the structural similarities between various industry models for portfolio credit risk management. For more information on copulas we refer to Appendix A and to Embrechts, McNeil, and Straumann (2001).

Proposition 3.2. *Let* $\left(X_i, (D_j^i)_{0 \le j \le n}\right)_{1 \le i \le m}$ *and* $\left(\widetilde{X}_i, (\widetilde{D}_j^i)_{0 \le j \le n}\right)_{1 \le i \le m}$ *be a pair of latent variable models with state vectors* $\mathbf{S}$ *and* $\widetilde{\mathbf{S}}$ *respectively. The models are equivalent if*

(i) The marginal distributions of the random vectors $\mathbf{S}$ *and* $\widetilde{\mathbf{S}}$ *coincide, i.e.*

$$P\left(X_i \le D_j^i\right) = P\left(\widetilde{X}_i \le \widetilde{D}_j^i\right),\ j \in \{0, \dots, n\},\ i \in \{1, \dots, m\}.$$

(ii) $\mathbf{X}$ *and* $\widetilde{\mathbf{X}}$ *admit the same copula.*

Note that in a model with only two states condition (i) simply means that the individual default probabilities $(\overline{p}_i)_{1 \le i \le m}$ are identical in both models. The converse of the result is not generally true. If two latent variable models are equivalent then $\mathbf{X}$ and $\widetilde{\mathbf{X}}$ need not necessarily have the same copula.

Proof. For notational simplicity consider the case $m = 2$. Denote by C the copula of $\mathbf{X}$ and recall the following identity (see (25) in Appendix A for more details).

$$P(X_1 \le x_1, X_2 \le x_2) = C(P(X_1 \le x_1), P(X_2 \le x_2)), \quad x_1, x_2 \in \mathbb{R}.$$

Write $u_{i,j} := P\left(X_i \le D_j^i\right) = P\left(\widetilde{X}_i \le \widetilde{D}_j^i\right),\ j \in \{0, \dots, n\},\ i = 1, 2$. Hence we get

$$\begin{aligned}
P(S_1 = j_1, S_2 = j_2) &= P\left(X_1 \in \left(D_{j_1-1}^1, D_{j_1}^1\right], X_2 \in \left(D_{j_2-1}^2, D_{j_2}^2\right]\right) \\
&= P\left(X_1 \le D_{j_1}^1, X_2 \le D_{j_2}^2\right) - P\left(X_1 \le D_{j_1-1}^1, X_2 \le D_{j_2}^2\right) \\
&\quad - P\left(X_1 \le D_{j_1}^1, X_2 \le D_{j_2-1}^2\right) + P\left(X_1 \le D_{j_1-1}^1, X_2 \le D_{j_2-1}^2\right) \\
&= C\left(u_{1,j_1}, u_{2,j_2}\right) - C\left(u_{1,j_1-1}, u_{2,j_2}\right) - C\left(u_{1,j_1}, u_{2,j_2-1}\right) + C\left(u_{1,j_1-1}, u_{2,j_2-1}\right) \\
&= \dots = P(\widetilde{S}_1 = j_1, \widetilde{S}_2 = j_2)\,.
\end{aligned}$$

For the case $m > 2$ the proof follows by an analogous argument from the following useful identity. For all $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$ with $a_i \le b_i$, $i = 1, \dots, m$

$$P(a_1 \le X_1 \le b_1, \dots, a_m \le X_m \le b_m) = \sum_{i_1=1}^{2} \cdots \sum_{i_m=1}^{2} (-1)^{i_1+\cdots+i_m} F(x_{1,i_1}, \dots, x_{m,i_m}),$$

where F denotes the df of $\mathbf{X}$, $x_{i,1} = a_i$ and $x_{i,2} = b_i$. □

We now discuss a number of popular industry models.

Example 3.3 (CreditMetrics and KMV model). Structurally these models are quite similar; they differ with respect to the approach used for calibrating individual default probabilities. In both models the latent vector $\mathbf{X}$ is assumed to have a multivariate normal distribution and X_i is interpreted as a change in asset value for obligor i over the time horizon of interest; D_0^i is chosen so that the probability of default for company i is the same as the historically observed default rate for companies of a similar credit quality. In CreditMetrics the classification of companies into groups of similar credit quality is generally based on an external rating system, such as that of Moodys or Standard & Poors; see RiskMetrics-Group (1997) for details. In KMV the so-called *distance-to-default* is used as state variable for credit quality. Essentially this quantity is computed using the Merton (1974) model for pricing defaultable securities, the main input being the value and volatility of a firm's equity; details can be found in KMV-Corporation (1997). In both models the covariance matrix Σ of $\mathbf{X}$ is calibrated using a factor model. It is assumed that the components of $\mathbf{X}$ can be written as

$$X_i = \sum_{j=1}^{p} a_{i,j}\Theta_j + \sigma_i \varepsilon_i + \mu_i, \quad i = 1, \dots, d\,, \tag{2}$$

for some $p < m$, a p-dimensional random vector $\mathbf{\Theta} \sim N_p(\mathbf{0}, \Omega)$ and independent standard normally distributed rv's $\varepsilon_1 \dots, \varepsilon_m$, which are also independent of $\mathbf{\Theta}$. Obviously, this implies that Σ is of the form $\Sigma = A\Omega A' + \mathrm{diag}(\sigma_1^2, \dots, \sigma_m^2)$. In practice the random vector $\mathbf{\Theta}$ represents country- and industry effects; calibration of the factor weights a_{ij} is achieved using "ad-hoc" economic arguments combined with statistical analysis of asset returns. Obviously, both models work with a Gaussian copula for the latent variable vector $\mathbf{X}$ and are hence structurally similar. In particular, by Proposition 3.2 the two-state versions of both models are equivalent, provided that the individual default probabilities $(\overline{p}_i)_{1 \le i \le m}$ are identical and that the correlation-matrix of $\mathbf{X}$ is the same in both models.

Example 3.4 (The model of Li (2001)). This model is quite popular with practitioners in pricing basket credit derivatives. Li interprets X_i as default-time of company i and assumes that X_i is exponentially distributed with parameter λ_i, i.e. $F_i(t) = 1 - \exp(-\lambda_i t)$. Obviously, company i has defaulted by time T if and only if $X_i \le T$, so that $\overline{p}_i = F_i(T)$. To determine the multivariate distribution of $\mathbf{X}$ Li assumes that $\mathbf{X}$ has the Gaussian copula C_R^{Ga} for some correlation matrix R (see for instance (26) in the Appendix), so that

$$P(X_1 \le t_1, \dots, X_m \le t_m) = C_R^{\mathrm{Ga}}(F_1(t_1), \dots, F_m(t_m))\,.$$

Again, this model is equivalent to a KMV-type model provided that individual default probabilities coincide and that the correlation matrix of the asset-value change $\mathbf{X}$ in the KMV-type model equals R.

While most latent variable models popular in industry are based on the Gaussian copula, there is no reason why we have assume a Gaussian copula. In fact, simulations presented in Section 5.1 show that the choice of copula may be very critical to the tail of the distribution of

M. We now give a theoretical explanation for this observation. For simplicity we restrict ourselves to two-state models. Consider a subgroup of k companies $\{i_1, \ldots, i_k\} \subset \{1, \ldots, m\}$, with individual default probabilities $\bar{p}_{i_1}, \ldots, \bar{p}_{i_k}$. Then

$$P(Y_{i_1} = 1, \ldots, Y_{i_k} = 1) = P\left(X_{i_1} \leq D_0^{i_1}, \ldots, X_{i_k} \leq D_0^{i_k}\right) = C_{i_1,\ldots,i_k}\left(\bar{p}_{i_1}, \ldots, \bar{p}_{i_k}\right), \quad (3)$$

where $C_{i_1,\ldots,i_k}$ denotes the corresponding k-dimensional margin of C. If $\mathbf{X}$ has an exchangeable copula (i.e. the copula of an exchangeable uniform random vector), and if all individual default probabilities are equal to some constant π, $\mathbf{Y}$ is exchangeable and (3) reduces to the useful formula $\pi_k = C_{1,\ldots,k}(\pi, \ldots, \pi)$, $1 \leq k \leq m$. It is obvious from these relations that the tail of M is largely affected by the tendency of the copula to produce small values in several margins simultaneously. This is related to the notion of lower tail dependence, which we now recall.

Definition 3.5 (Lower Tail Dependence). Given rv's X_1 and X_2 with continuous marginal distributions F_1 and F_2 and hence with unique copula C. The coefficient of lower tail dependence is defined to be $\lambda_\ell = \lim_{u \to 0} \frac{C(u,u)}{u}$, provided the limit exists. If $\lambda_\ell \in (0, 1]$ then the pair (X_1, X_2) (or the copula C) is called lower tail-dependent; if $\lambda_\ell = 0$, the pair (X_1, X_2) is called asymptotically independent.

Denote by $F_i^{\leftarrow}$ the quantile function of F_i, $i = 1, 2$. Again since $\mathbf{X}$ has continuous margins, we obtain $\lambda_\ell = \lim_{u \to 0} P\left(X_2 \leq F_2^{\leftarrow}(u) \mid X_1 \leq F_1^{\leftarrow}(u)\right)$, i.e. λ_ℓ gives the limiting conditional probability that X_2 lies below its u-quantile, given that X_1 lies below its u-quantile (or vice versa). It is well-known that for bivariate Gaussian random variables with correlation $\rho < 1$ (and hence the Gaussian copula with parameter $\rho < 1$) λ_ℓ is zero – see for instance Embrechts, McNeil, and Straumann (2001) – such that models based on a Gaussian copula might underestimate the probability of many joint defaults.

3.2 Latent variable models with non-Gaussian dependence structure

The KMV/CreditMetrics-type models can accommodate a wide range of different correlation structures for the latent variables. This is clearly an advantage in modelling a portfolio where obligors are exposed to several risk factors and where the exposure to different risk factors differs markedly accross obligors such as a portfolio of loans to companies from different industries or countries. The following class of models preserves this flexibility of the KMV/CreditMetrics-type models, but allows us to work with copulas which are (lower) tail dependent.

Example 3.6 (Normal mean-variance mixtures). In this class we start with an m-dimensional vector $\mathbf{Z} \sim N(\mathbf{0}, \Sigma)$ and some rv W which is independent of Z. The latent variables are then of the form

$$X_i := \mu_i(W) + g(W) Z_i, \quad 1 \leq i \leq m, \quad (4)$$

for functions $\mu_i : \mathbb{R} \to \mathbb{R}$ and $g_i : \mathbb{R} \to (0, \infty)$. In case that μ_i is independent of W for all i model (4) is called a normal variance mixture.

Examples include the multivariate t distribution with mean $\boldsymbol{\mu}$ and degrees of freedom ν, denoted by $t_m(\nu, \boldsymbol{\mu}, \Sigma)$, and the generalized hyperbolic distribution. To obtain $t_m(\nu, \mathbf{0}, \Sigma)$-distributed latent variables we put

$$\mu_i \equiv 0, \ i = 1, \ldots, m, \quad g(w) = \sqrt{\frac{\nu}{w}}, \text{ and } W \sim \chi^2(\nu); \quad (5)$$

in that case $\mathbf{X}$ has standard t distributed marginals and, for $\nu \geq 2$, covariance matrix $\frac{\nu}{\nu-2}\Sigma$. It is well-known that the t copula is tail-dependent; see for instance Section 4 of Embrechts, McNeil, and Straumann (2001) for details. To obtain a generalized hyperbolic distribution we assume that the mixing variable W follows a normal inverse Gaussian distribution and take $\mu_i(W) = \beta_i W^2$ for constants β_i and $g(W) = W$. This distribution is a general mean-variance mixture. The generalized hyperbolic distribution has been advocated as a model for univariate stock returns by Eberlein and Keller (1995).

In a normal mean-variance mixture model the default condition may be written as

$$X_i \le D_0^i \iff Z_i \le D_0^i h_1(W) - h_{i,2}(W) =: \widetilde{D}_0^i \,, \tag{6}$$

where $h_1(w) = 1/g(w)$ and $h_{i,2}(w) = \mu_i(w)/g(w)$. This suggests the following economic interpretation: Z_i and D_0^i represent the asset value respectively an a-priori estimate of the liabilities of company i; default occurs if the asset value lies below the actual default threshold $\widetilde{D}_0^i$, which is obtained by applying a multiplicative and an additive shock to the a-priori estimate D_0^i. If we interpret this shock as a stylized representation of global factors such as the overall liquidity and risk appetite in the banking system, it makes sense to assume that for all obligors these shocks are driven by the same rv W.

In this paper we are particularly interested in normal variance mixtures such as multivariate t. In this class of models the correlation matrix of $\mathbf{X}$ and $\mathbf{Z}$ coincide, provided of course that $E(g(W)) < \infty$. Moreover, if $\boldsymbol{Z}$ follows the linear factor model (2), $\mathbf{X}$ inherits the linear factor structure from $\mathbf{Z}$. Note however, that the "systematic factors" $g(W)\mathbf{\Theta}$ and the "idiosyncratic factors" $g(W)\sigma_i\varepsilon_i$, $1 \le i \le m$, are no longer independent but merely uncorrelated.

Alternatively we could use parametric copulas in closed-form to model the distribution of $\mathbf{X}$. An example is provided by the class of so-called Archimedian copulas.

Example 3.7 (Archimedian copulas). A k-dimensional Archimedean copula is the distribution function of an exchangeable uniform random vector and has the form

$$C_{1,\dots,k}(u_1, \dots, u_k) = \phi^{-1}\left(\phi(u_1) + \cdots + \phi(u_k)\right), \tag{7}$$

where $\phi : [0,1] \mapsto [0,\infty]$ is a continuous, strictly decreasing function, known as the copula *generator* which satisfies $\phi(0) = \infty$ and $\phi(1) = 0$; ϕ^{-1} is the generator inverse. A theorem of Kimberling (1974) (see also Schweizer and Sklar (1983)) shows that a necessary and sufficient condition for (7) to define a proper copula for all k is that ϕ^{-1} is a *completely monotonic* function on $[0,\infty)$, i.e. $(-1)^k \frac{d^k}{dt^k}\phi^{-1}(t) \ge 0, \quad k \in \mathbb{N}$. There are many possibilities for generating Archimedean copulas (Nelsen 1999). A useful class of Archimedian copulas with lower tail dependence is Clayton's copula family, which is obtained by taking the generator $\phi_\theta(t) = t^{-\theta} - 1$; it may be verified that the generator inverse is a completely monotonic function. Archimedian copulas suffer from the deficiency that they are not rich in parameters and cannot model a fully flexible dependence between the latent variables. Nonetheless they yield useful parsimonious models for relatively small homogeneous portfolios, which are easy to simulate from; see Example 4.14 below.

There are various other methods of constructing general m-dimensional copulas; useful references are Joe (1997), Nelsen (1999) and Lindskog (2000).

4 Mixture models

In a mixture model the default probability of an obligor is assumed to depend on a set of common economic factors such as macroeconomic variables; given the default probabilities defaults of different obligors are independent. Dependence between defaults hence stems from the dependence of the default-probabilities on a set of common factors.

Definition 4.1 (Bernoulli Mixture Model). Given some $p < m$ and a p-dimensional random vector $\mathbf{\Psi} = (\Psi_1, \ldots, \Psi_p)$, the random vector $\mathbf{Y} = (Y_1, \ldots, Y_m)'$ follows a Bernoulli mixture model with factor vector $\mathbf{\Psi}$, if there are functions $Q_i : \mathbb{R}^p \to [0,1]$, $1 \le i \le m$, such that conditional on $\mathbf{\Psi}$ the default indicator $\mathbf{Y}$ is a vector of independent Bernoulli random variables with $P(Y_i = 1|\mathbf{\Psi}) = Q_i(\mathbf{\Psi})$.

For $\mathbf{y} = (y_1, \ldots, y_m)'$ in $\{0,1\}^m$ we have that

$$P(\mathbf{Y} = \mathbf{y} \mid \mathbf{\Psi}) = \prod_{i=1}^{m} Q_i(\mathbf{\Psi})^{y_i}(1 - Q_i(\mathbf{\Psi}))^{1-y_i}, \tag{8}$$

and the unconditional distribution of the default indicator vector $\mathbf{Y}$ is obtained by integrating over the distribution of the factor vector $\mathbf{\Psi}$.

Example 4.2 (CreditRisk$^+$). CreditRisk$^+$ may be represented as a Bernoulli mixture model where the distribution of the default indicators is given by

$$P(Y_i = 1 \mid \mathbf{\Psi}) = Q_i(\mathbf{\Psi}) \text{ for } Q_i(\mathbf{\Psi}) = 1 - \exp(-\mathbf{w}_i'\mathbf{\Psi}). \tag{9}$$

Here $\mathbf{\Psi} = (\Psi_1, \ldots, \Psi_p)'$ is a vector of independent gamma distributed macroeconomic factors with $p < m$ and $\mathbf{w}_i = (w_{i,1}, \ldots, w_{i,p})'$ is a vector of positive, constant factor weights.

We note that CreditRisk$^+$ is usually presented as a *Poisson* mixture model. In this more common presentation it is assumed that, conditional on $\mathbf{\Psi}$, the default of counterparty i occurs independently of other counterparties with a Poisson intensity given by $\Lambda_i(\mathbf{\Psi}) = \mathbf{w}_i'\mathbf{\Psi}$. Although this assumption makes it possible to default more than once, a realistic model calibration generally ensures that the probability of this happening is very small. The conditional probability given $\mathbf{\Psi}$ that a counterparty defaults over the time period of interest (whether once or more than once) is given by

$$1 - \exp(-\Lambda_i(\mathbf{\Psi})) = 1 - \exp(-\mathbf{w}_i'\mathbf{\Psi}),$$

so that we obtain the Bernoulli mixture model in (9). The Poisson formulation of CreditRisk$^+$ has the pleasant analytical feature that the distribution of the number of defaults in the portfolio is equal to the distribution of a sum of independent negative binomial random variables, as is shown in Gordy (2000). For more details on CreditRisk$^+$ and its calibration in practice see Credit-Suisse-Financial-Products (1997).

A similar argument shows that the Cox-process models of Lando (1998) or Duffie and Singleton (1999) also lead to Bernoulli-mixture models for the default indicator at a given time T.

4.1 One-factor Bernoulli mixture models

In many practical situations it is useful to consider a one-factor model. The information may not always be available to calibrate a model with more factors, and one-factor models may be fitted statistically to default data without great difficulty, as is shown in Section 5.2. Their behaviour for large portfolios is also particularly easy to understand using results in Section 4.2.

Throughout this section Ψ is a rv with values in $\mathbb{R}$ and $Q_i(\Psi) : \mathbb{R} \to [0,1]$ a set of functions such that, conditional on Ψ, the default indicator $\mathbf{Y}$ is a vector of independent Bernoulli random variables with $P(Y_i = 1|\Psi) = Q_i(\Psi)$. We now consider a variety of special cases.

Exchangeable Bernoulli mixture models. A further simplification occurs in the case that the functions Q_i are all identical. In this case the Bernoulli-mixture model is termed *exchangeable* since the random vector $\mathbf{Y}$ is exchangeable. It is convenient to introduce the rv $Q := Q_1(\Psi)$ and to denote the df of this mixing variable by $G(q)$. The distribution of the number of defaults M in this model is given by

$$P(M = k) = \binom{m}{k} \int_0^1 q^k (1-q)^{m-k} dG(q) \,. \tag{10}$$

Further simple calculations give $\pi = E(Y_1) = E\left(E(Y_1 \mid Q)\right) = E(Q)$ and, more generally,

$$\pi_k \;\; = \;\; P(Y_1 = 1, \dots, Y_k = 1) = E\left(E(Y_1 \cdots Y_k \mid Q)\right) = E(Q^k), \tag{11}$$

so that unconditional default probabilities of first and higher order are seen to be moments of the mixing distribution. Moreover, for $i \neq j$

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

which means that in an exchangeable Bernoulli mixture model the default correlation ρ_Y defined in (1) is always nonnegative. Any value of ρ_Y in $[0,1]$ can be obtained by an appropriate choice of the mixing distribution G. In particular, if $\rho_Y = \operatorname{var}(Q) = 0$ the rv Q has a degenerate distribution with all mass concentrated on the point π and the default indicators are independent. The case $\rho_Y = 1$ corresponds to a model where $\pi = \pi_2$ and the distribution of Q is concentrated on the points 0 and 1.

The following exchangeable Bernoulli mixture models are frequently used in practice.

- Beta mixing-distribution. Here $Q \sim \text{Beta}(a,b)$ with density $g(q) = \beta(a,b)^{-1} q^{a-1} (1-q)^{b-1}$, $a, b > 0$, where β denotes the beta function. This model is much the same as a one-factor exchangeable version of CreditRisk$^+$, as is shown in Frey and McNeil (2002).

- Probit-normal mixing-distribution. Here $Q = \Phi(\mu + \sigma\Psi)$ for $\Psi \sim N(0,1)$, $\mu \in \mathbb{R}$, $\sigma > 0$ and Φ the standard normal df. It turns out that this model can be viewed as a one-factor version of the CreditMetrics and KMV-type models; this is a special case of a general result in Section 4.3.

- Logit-normal mixing-distribution. Here $Q = 1/(1 + \exp(\mu + \sigma\Psi))$ for $\Psi \sim N(0,1)$, $\mu \in \mathbb{R}$ and $\sigma > 0$. This model can be thought of as a one-factor version of the CreditPortfolioView model of Wilson (1997); see Section 5 of Crouhy, Galai, and Mark (2000) for details.

In the model with beta mixing distribution the higher order default probabilities π_k and the distribution of M can be computed explicitly; see Frey and McNeil (2001). Calculations for the logit-normal, probit-normal and other models generally require numerical evaluation of the integrals in (10) and (11). If we fix any two of π, π_2 or ρ_Y in a beta, logit-normal or probit-normal model, then this fixes the parameters μ and σ of the mixing distribution and higher order joint default probabilities are automatic.

Bernoulli regression models. These models are quite useful for practical purposes. In Bernoulli regression models deterministic covariates are allowed to influence the probability of default; the effective dimension of the mixing distribution is still one. The individual conditional default probabilities are now of the form

$$Q_i(\Psi) = Q(\Psi, \mathbf{z_i}) \,, \; 1 \leq i \leq m,$$

where $\mathbf{z}_i \in \mathbb{R}^k$ is a vector of deterministic covariates and $Q : \mathbb{R}\times\mathbb{R}^k \to [0,1]$ is strictly increasing in its first argument. There are many possibilities for this function and a particularly tractible specification is

$$Q(\Psi, \mathbf{z}_i) = h(\boldsymbol{\sigma}'\mathbf{z}_i\Psi + \boldsymbol{\mu}'\mathbf{z}_i)\,, \tag{12}$$

where $h : \mathbb{R} \to [0,1]$ is some strictly increasing *link function*, such as $h(x) = \Phi(x)$ or $h(x) = (1+\exp(-x))^{-1}$; $\boldsymbol{\mu} = (\mu_1 \ldots, \mu_k)'$ and $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_k)'$ are vectors of regression parameters and $\boldsymbol{\sigma}'\mathbf{z}_i > 0$. If Ψ is taken to be a standard normally distributed factor then with the above choices of link functions we have a probit-normal or logit-normal mixture distribution for every obligor. For alternative specifications to (12) for the form of the regression relationship see for instance Joe (1997), page 216.

Obviously if $\mathbf{z}_i = \mathbf{z}$, $\forall i$, so that all risks have the same covariates, then we are back in the situation of full exchangeability. Note also that, since the function $Q(\psi, \cdot)$ is increasing in ψ, the conditional default probabilities form a comonotonic random vector; in particular, in a state of the world where the default-probability is high for one counterparty it is high for all counterparties. This is a useful feature for modelling default-probabilities corresponding to different rating classes.

Example 4.3 (Model for several exchangeable groups). The regression structure includes partially exchangeable models where we define a number of groups within which risks are exchangeable; these might represent rating classes according to some internal or rating agency classification.

Assume we have k groups and $r(i) \in \{1, \ldots, k\}$ gives the group membership of individual i. Assume further that the vectors $\mathbf{z}_i$ are k-dimensional unit vectors of the form $\mathbf{z}_i = \mathbf{e}_{r(i)}$ so that $\boldsymbol{\sigma}'\mathbf{z}_i = \sigma_{r(i)}$ and $\boldsymbol{\mu}'\mathbf{z}_i = \mu_{r(i)}$. If we use construction (12) above then for an individual i we have

$$Q_i(\Psi) = h(\mu_{r(i)} + \sigma_{r(i)}\Psi), \tag{13}$$

where $\sigma_{r(i)} > 0$. Inserting this specification in (8) we can find the conditional distribution of the default indicator vector. Suppose there are m_r individual in group r for $r = 1, \ldots, k$ and write M_r for the number of defaults. The conditional distribution of the vector $\mathbf{M} = (M_1, \ldots, M_k)'$ is given by

$$P(\mathbf{M} = \mathbf{l} \mid \Psi) = \prod_{r=1}^{k} \binom{m_r}{l_r} \left(h(\mu_r + \sigma_r\Psi)\right)^{l_r} \left(1 - h(\mu_r + \sigma_r\Psi)\right)^{m_r - l_r}, \tag{14}$$

where $\mathbf{l} = (l_1, \ldots, l_k)'$. A model of precisely the form (14) will be fitted to Standard and Poor's default data in Section 5.2. The asymptotic behaviour of such a model (when m is large) is investigated in Example 4.10.

4.2 Loss distributions for large portfolios in Bernoulli mixture models

We now provide some asymptotic results for large portfolios in Bernoulli mixture models. These results will be useful for the analysis of model risk in mixture models and for the determination of moments of the number of defaults.

We use the following setup: $(e_i)_{i\in\mathbb{N}}$ is a sequence of positive constants representing exposures; $(\Delta_i)_{i\in\mathbb{N}}$ is a sequence of rv's with values in $(0,1]$ representing percentages losses given that default occurs; $(Y_i)_{i\in\mathbb{N}}$ is a sequence of default indicators as usual. Define $L_i = e_i\Delta_i Y_i$ to be the loss of company i (if any) and $L^{(m)} = \sum_{i=1}^{m} L_i$ to be the loss for a portfolio of size m. We make the following assumptions.

Assumption 4.4.

(i) There is a p-dimensional random vector $\mathbf{\Psi}$ and functions ℓ_i: $\text{supp}(\mathbf{\Psi}) \to [0,1]$ such that conditional on $\mathbf{\Psi}$ the $(L_i)_{i\in\mathbb{N}}$ form a sequence of independent rv's with mean $\ell_i(\boldsymbol{\psi}) = E(L_i \mid \mathbf{\Psi} = \boldsymbol{\psi})$.

(ii) There is a function $\overline{\ell} : \text{supp}(\mathbf{\Psi}) \to \mathbb{R}^+$such that

$$\lim_{m\to\infty} \frac{1}{m} E\left(L^{(m)} \mid \mathbf{\Psi} = \boldsymbol{\psi}\right) = \lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} \ell_i(\boldsymbol{\psi}) = \overline{\ell}(\boldsymbol{\psi})$$

for all $\boldsymbol{\psi} \in \text{supp}(\mathbf{\Psi})$. We refer to $\overline{\ell}(\boldsymbol{\psi})$ as the asymptotic conditional loss function.

(iii) There is some $C < \infty$ such that $\sum_{i=1}^{m} (e_i/i)^2 < C$ for all m, so that the exposures are bounded in this sense.

Proposition 4.5. *Given a sequence $L^{(m)} = \sum_{i=1}^{m} L_i$ satisfying Assumption 4.4. Denote by $P(\cdot \mid \mathbf{\Psi} = \boldsymbol{\psi})$ the conditional distribution of the sequence $(L_i)_{i\in\mathbb{N}}$ given $\mathbf{\Psi} = \boldsymbol{\psi}$. Then*

$$\lim_{m\to\infty} \frac{1}{m} L^{(m)} = \overline{\ell}(\boldsymbol{\psi}) \quad P(\cdot \mid \mathbf{\Psi} = \boldsymbol{\psi})\, a.s. \text{ for all } \boldsymbol{\psi} \in \text{supp}(\mathbf{\Psi})\,.$$

Proof. Our proof is based on the following Lemma (Petrov (1975), Theorem IX.12).

Lemma 4.6. *Let $(Z_i)_{i\in\mathbb{N}}$ be a sequence of independent rv's with $E(Z_i) = 0$. Suppose that $\sum_{i=1}^{\infty} E(|Z_i|^2)/i^2 < \infty$. Then $\frac{1}{m}\sum_{i=1}^{m} Z_i \to 0$ a.s.*

To apply the lemma we put $Z_i := L_i - \ell_i(\boldsymbol{\psi})$. Since $0 \le e_i \le 1$ we get $\sum_{i=1}^{\infty} E(Z_i^2)/i^2 \le \sum_{i=1}^{\infty} (e_i/i)^2$ which is finite by Assumption 4.4(iii). □

Remark 4.7. 1) A related result has independently been obtained by Gordy (2001).
2) If we put $\Delta_i = e_i \equiv 1$, Proposition 4.5 applies to $M^{(m)} = \sum_{i=1}^{m} Y_i$. For a given sequence $(Y_i)_{i\in\mathbb{N}}$ following a p-factor Bernoulli mixture model with default probabilities $Q_i(\boldsymbol{\psi})$ Assumption 4.4 reduces to

$$\lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} Q_i(\boldsymbol{\psi}) = \overline{Q}(\boldsymbol{\psi}) \text{ for some function } \overline{Q} :\ \text{supp}(\mathbf{\Psi}) \to [0,1]\,. \tag{15}$$

Suppose that (15) holds. Since $\frac{1}{m} M^{(m)}$ is bounded, we obtain by dominated convergence $\lim_{m\to\infty} E(|f(\frac{1}{m}M^{(m)}) - f(\overline{Q}(\mathbf{\Psi}))|) = 0$ for every continuous $f : [0,1] \to \mathbb{R}$. This can be useful for computing the moments of $M^{(m)}$ for large portfolios. We have for $k \in \mathbb{N}$

$$E\left((M^{(m)})^k\right) = m^k E\left(\left(\frac{1}{m} M^{(m)}\right)^k\right) \approx m^k E\left(\overline{Q}(\mathbf{\Psi})^k\right)\,. \tag{16}$$

3) In standard portfolio credit risk models it is assumed that default indicators and loss given default are independent. While this assumption is highly debatable, as one would expect that Δ_i should generally be higher in bad states of the world (represented by values of $\mathbf{\Psi}$ leading to high default probabilities) than in good ones, it makes more sense to assume that $(Y_i)_{i\in\mathbb{N}}$ and $(\Delta_i)_{i\in\mathbb{N}}$ are conditionally independent given $\mathbf{\Psi}$. In that case we have $\ell_i(\boldsymbol{\psi}) = e_i \Delta_i(\boldsymbol{\psi}) Q_i(\boldsymbol{\psi})$, where $\Delta_i(\boldsymbol{\psi})$ represents the mean loss given default of company i given $\mathbf{\Psi} = \boldsymbol{\psi}$. If we moreover assume that conditional default probabilities and conditional mean exposure are independent and that $\overline{\Delta}(\boldsymbol{\psi}) = \lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^{m} e_i \Delta_i(\boldsymbol{\psi})$ exists, the asymptotic loss function is of the multiplicative form $\overline{\ell}(\boldsymbol{\psi}) = \overline{\Delta}(\boldsymbol{\psi})\overline{Q}(\boldsymbol{\psi})$; we will work with this form of asymptotic conditional loss function $\overline{\ell}$ in Section 5.1.

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.

Proposition 4.8. *Consider a sequence* $L^{(m)} = \sum_{i=1}^{m} L_i$ *satisfying Assumption 4.4 with a one-dimensional mixing variable* Ψ *with df* $G(\psi)$. *Assume that the conditional asymptotic loss function* $\overline{\ell}(\psi)$ *is strictly increasing and right 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*

$$\lim_{m\to\infty} \frac{1}{m}\, q_\alpha(L^{(m)}) = \overline{\ell}(q_\alpha(\Psi))\,. \tag{17}$$

Remark 4.9. 1) The assumption that $\overline{\ell}$ is strictly increasing makes sense if we assume that low (high) values of Ψ correspond to good (bad) states of the world with conditional default probabilities and losses given default lower (higher) than average.
2) Consider two exchangeable Bernoulli mixture models with mixing distributions $G_i(q) = P(Q_i < q)$, $i = 1, 2$. Suppose that the tail of G_1 is heavier than the tail of G_2, i.e. that we have $G_1(q) < G_2(q)$ for q close to 1. Then Proposition 4.8 implies that for large m the tail of $M^{(m)}$ is heavier in model 1 than in model 2.
3) Proposition 4.8 has implications for setting capital adequacy rules for loan books. This point is discussed in Gordy (2001).

Proof. Using the structure of the Bernoulli mixture model, Fatou's Lemma and Proposition 4.5 we obtain for any $\varepsilon > 0$

$$\begin{aligned}\limsup_{m\to\infty} P\left(L^{(m)} \le m(\overline{\ell}(q_\alpha(\Psi)) - \varepsilon)\right) &\le \int_{\mathbb{R}} \limsup_{m\to\infty} P\left(L^m \le m(\overline{\ell}(q_\alpha(\Psi)) - \varepsilon) | \Psi = \psi\right) dG(\psi)\\ &\le \int_{\mathbb{R}} 1_{\{\overline{\ell}(\psi) < \overline{\ell}(q_\alpha(\Psi)) - \varepsilon/2\}}\, dG(\psi).\end{aligned}$$

Since $\overline{\ell}$ is strictly increasing, there is some $\delta > 0$ such that the last integral is no larger than $G(q_\alpha(\Psi) - \delta)$, which is smaller than α by definition of $q_\alpha(\Psi)$.

Similarly we have

$$\liminf_{m\to\infty} P\left(L^{(m)} \le m(\overline{\ell}(q_\alpha(\Psi)) + \varepsilon)\right) \ge \int_{\mathbb{R}} 1_{\{\overline{\ell}(\psi) < \overline{\ell}(q_\alpha(\Psi)) + \varepsilon/2\}}\, dG(\psi)\,.$$

Since $\overline{\ell}$ is increasing and right continuous, there is some $\delta > 0$ such that the last integral is no larger than $G(q_\alpha(\Psi) + \delta)$, which is strictly larger than α by assumption. Hence for m large we have $m(\overline{\ell}(q_\alpha(\Psi)) - \varepsilon) \le q_\alpha(L^{(m)}) \le m(\overline{\ell}(q_\alpha(\Psi)) + \varepsilon)$. ☐

Example 4.10. Consider the Bernoulli regression model for k exchangeable groups defined by (13). The assumption implied by Equation (15) translates to

$$\lim_{m\to\infty} \frac{1}{m} \sum_{r=1}^{k} m_r^{(m)} h(\mu_r + \sigma_r \psi) = \overline{Q}(\psi),$$

for some function $\overline{Q}$, which is fulfilled if $m_r^{(m)}/m$, the proportions of obligors in each group, converge to fixed constants λ_r as $m \to \infty$. Assuming unit exposures and 100% losses given default our asymptotic conditional loss function is $\overline{\ell}(\psi) = \sum_{r=1}^{k} \lambda_r h(\mu_r + \sigma_r \psi)$. Since Ψ has a standard normal distribution (17) implies for large m

$$q_\alpha(L^{(m)}) \approx m \sum_{r=1}^{k} \lambda_r h(\mu_r + \sigma_r \Phi^{-1}(\alpha))\,.$$

4.3 Relation to latent variable models

At a first glance latent variable models and Bernoulli mixture models appear to be very different types of models. However, as has already been observed by Gordy (2000), these differences are often related more to presentation and interpretation than to mathematical substance. In this section we provide a fairly general result linking latent variable models and mixture models. Results on the relationship between latent variable models and mixture models are useful from a theoretical and an applied perspective. From a theoretical viewpoint results on the connection between these model classes help to distinguish essential from inessential features of credit risk models; from a practical point of view a link between the different types of models enables us to apply numerical and statistical techniques for solving and calibrating the models, which are natural in the context of mixture models, also to latent variable models and vice versa. We will make frequent use of this in Section 5.

The following condition ensures that a latent variable model can be written as a Bernoulli mixture model.

Definition 4.11. A latent-variable-vector $\mathbf{X}$ has a p-dimensional conditional independence structure with conditioning variable $\mathbf{\Psi}$, if there is some $p < m$ and a p-dimensional random vector $\mathbf{\Psi} = (\Psi_1, \ldots, \Psi_p)'$ such that conditional on $\mathbf{\Psi}$ the rv's $(X_i)_{1 \le i \le m}$ are independent.

Proposition 4.12. *Consider an m-dimensional latent variable vector $\mathbf{X}$ and a p-dimensional $(p < m)$ random vector $\mathbf{\Psi}$. Then the following are equivalent.*

(i) $\mathbf{X}$ has p-dimensional conditional independence structure with conditioning variable $\mathbf{\Psi}$.

(ii) For any choice of thresholds D_0^i, $1 \le i \le m$ the default indicators $Y_i = 1_{\{X_i \le D_0^i\}}$ follow a Bernoulli mixture model with factor $\mathbf{\Psi}$; the conditional default probabilities are given by $Q_i(\mathbf{\Psi}) = P(X_i \le D_0^i \mid \mathbf{\Psi})$.

Proof. Suppose that (i) holds. Define for $\mathbf{y} \in \{0,1\}^m$ the set $A := \{1 \le i \le m : y_i = 1\}$ and let $A^c = \{1, \ldots, m\} - A$. We have

$$\begin{aligned} P(\mathbf{Y} = \mathbf{y} \mid \mathbf{\Psi}) &= P\left(\bigcap_{i \in A} \{X_i \le D_0^i\} \bigcap_{i \in A^c} \{X_i > D_0^i\} \mid \mathbf{\Psi} \right) \\ &= \prod_{i \in A} P(X_i \le D_0^i \mid \mathbf{\Psi}) \prod_{i \in A^c} (1 - P(X_i \le D_0^i \mid \mathbf{\Psi})\,. \end{aligned}$$

Hence conditional on $\mathbf{\Psi}$ the Y_i are independent Bernoulli variates with success-probability $Q_i(\mathbf{\Psi}) := P(X_i \le D_0^i \mid \mathbf{\Psi})$. The converse is obvious. □

Example 4.13 (Normal mean-variance mixtures with factor structure). Suppose that $\mathbf{X} = \boldsymbol{\mu}(W) + g(W)\mathbf{Z}$ for W independent of $\mathbf{Z}$ and that $\mathbf{Z}$ follows the linear factor model (2), i.e. $Z_i = \sum_{j=1}^p a_{i,j}\Theta_j + \sigma_i \varepsilon_i$ for a random vector $\mathbf{\Theta} \sim N_p(\mathbf{0}, \Omega)$ and independent $N(0,1)$-distributed rv's $\varepsilon_1 \ldots, \varepsilon_m$ independent of $\mathbf{\Theta}$. Then $\mathbf{X}$ has a $(p+1)$-dimensional conditional independence structure. Define the $(p+1)$-dimensional random vector $\mathbf{\Psi}$ by $\mathbf{\Psi} := (\Theta_1, \ldots, \Theta_p, W)'$. Conditional on $\mathbf{\Psi}$ the rv's X_i are obviously independent and normally distributed with mean $\mu_i(W) + g(W)\sum_{j=1}^p a_{ij}\Theta_j$ and variance $(g(W)\sigma_i)^2$. The equivalent Bernoulli mixture model is now easy to compute. Given thresholds $(D_0^i)_{1 \le i \le m}$ we get that

$$Q_i(\mathbf{\Psi}) = P(X_i < D_0^i \mid \mathbf{\Psi}) = \Phi\left(\frac{D_0^i - \mu_i(W) - g(W)\sum_{j=1}^p a_{ij}\Theta_j}{g(W)\sigma_i} \right). \tag{18}$$

In the special case of multivariate t latent variables we obtain

$$Q_i(\mathbf{\Psi}) = \Phi\left(\sigma_i^{-1} \left(D_0^i \sqrt{W/\nu} - \sum\nolimits_{j=1}^p a_{ij}\Theta_j \right) \right). \tag{19}$$

Example 4.14 (Archimedian copulas). As shown in the following lemma, which is essentially due to Marshall and Olkien (1988), latent variable models based on exchangeable Archimedian copulas possess a one-dimensional conditional independence structure.

Lemma 4.15. *Given a df F on $\mathbb{R}^+$ with Laplace transform $\varphi(x) = \int_0^\infty \exp(-xy)dF(y)$, and suppose that $F(0) = 0$. Denote by φ^{-1} the functional inverse of φ. Consider a rv $\Psi \sim F$ and a sequence $(U_i)_{1\le i\le m}$ of rv's which are conditionally independent given Ψ with conditional df $P(U_i \le u \mid \Psi = \psi) = \exp(-\psi\varphi^{-1}(u))$ for $u \in [0,1]$. Then*

$$P(U_1 \le u_1, \dots, U_m \le u_m) = \varphi(\varphi^{-1}(u_1), \dots, \varphi^{-1}(u_m)),$$

i.e. $(U_i)_{1\le i\le m}$ has an Archimedian copula with generator $\phi = \varphi^{-1}$. Moreover, every Archimedian copula can be obtained that way.

Proof. We have

$$\begin{aligned} P(U_1 \le u_1, \dots, U_n \le u_n) &= \int_0^\infty P(U_1 \le u_1, \dots, U_m \le u_m \mid \Psi = \psi)dF(\psi) \\ &= \int_0^\infty \exp\left(-\psi(\varphi^{-1}(u_1) + \ldots + \varphi^{-1}(u_m))\right) dF(\psi) \\ &= \varphi\left(\varphi^{-1}(u_1) + \ldots + \varphi^{-1}(u_m)\right) . \end{aligned}$$

Moreover, as shown in Joe (1997), every generator of an Archimedian copula is of the form $\phi = \varphi^{-1}$, where φ is the Laplace transform of some df $\tilde{F}$ on $\mathbb{R}^+$ with $\tilde{F}(0) = 0$. □

The Lemma gives an obvious recipe for simulating from an Archimedian copula with generator ϕ, provided that we know a df with Laplace transform equal to ϕ^{-1} so that we can simulate values of Ψ. For instance, in order to simuate from a Clayton copula we have to use a rv Ψ, which is gamma distributed.

Consider now a latent variable model (X_i, D_0^i), $1 \le i \le m$ where $\mathbf{X}$ has an exchangeable Archimedean copula with generator ϕ. Put $Y_i = 1_{\{X_i \le D_0^i\}}$ and $\overline{p}_i = P(Y_i = 1)$. Using Lemma 4.15, an equivalent Bernoulli mixture model is now straightforward to compute. Observe that for Ψ, $(U_i)_{1\le i\le m}$ as in the Lemma $\left(X_i, D_0^i\right)_{1\le i\le m}$ and $(U_i, \overline{p}_i)_{1\le i\le m}$ are two equivalent latent variable modes by Proposition 3.2. Moreover, the $(U_i)_{1\le i\le m}$ are obviously independent given Ψ and we obtain for the conditional default probabilities

$$P(U_i \le \overline{p}_i \mid \Psi = \psi) = Q_i(\psi) := \exp(-\psi\varphi^{-1}(\overline{p}_i)) .$$

A similar result was obtained independently in Schönbucher (2002).

Remark 4.16. The study of mixture representations for sequences of exchangeable Bernoulli random variables is related a the well-known result of De Finetti, which concerns *infinite* sequences of exchangeable Bernoulli random variables. A sequence $Y_1, Y_2, \dots$ is said to be exchangeable if the random vectors $(Y_1, \dots, Y_k)$ are exchangeable for all $k \in \mathbb{N}$. De Finetti proved that for any infinite sequence $Y_1, Y_2, \dots$ one can find a probability distribution G on $[0,1]$ such that for all $k \le m \in \mathbb{N}$

$$P(Y_1 = 1, \dots, Y_k = 1, Y_{k+1} = 0, \dots, Y_m = 0) = \int_0^1 q^k (1-q)^{m-k} dG(q). \tag{20}$$

This shows that any exchangeable model for $\mathbf{Y}$, which can be extended to arbitrary portfolio size m, has a representation as an exchangeable Bernoulli-mixture model.

5 Model risk and model calibration

Due to the complexity of the phenomena involved and the scarcity of reliable data, the specification of a portfolio credit risk model is a difficult task. In this section we attempt first to understand the sensitivity of the loss distribution with respect to the misspecification of parts of the model and/or key input parameters. It suffices here to simplify to a model for a homogeneous group.

We then discuss the calibration of portfolio default models in the following two situations.

- A large portfolio divided into rating categories with no other information pertinent to default. Here we look at the statistical fitting of a one-factor Bernoulli mixture model with regression structure to historical default data.
- A portfolio where the default potential of individual obligors is considered to be better understood so that they are treated more heterogeneously. Here we look at how the calibration of latent variable models using the approach of KMV/CreditMetrics may be improved using normal variance mixture models and historical data.

5.1 Model risk in a homogeneous group model

The impact of the copula choice in latent variable models. Since most latent variable models used in industry work with the Gaussian copula we are particularly interested in the sensitivity of the distribution of M wrt. the assumption of Gaussian dependence. We compare two models; a model with multivariate normal latent variables and a model where latent variables are multivariate t. For simplicity we consider a homogeneous group model i.e. we model $\mathbf{Z}$ by $Z_i = \sqrt{\rho}\Theta + \sqrt{1-\rho}\varepsilon_i$, $i = 1, \dots, m$ and $\rho \geq 0$. In the t case we put $X_i = \sqrt{\nu/W} Z_i$ for a rv $W \sim \chi^2(\nu)$ independent of $\mathbf{Z}$; in the normal case, which corresponds to $\nu = \infty$, we put $X_i = Z_i$, $i = 1, \dots, m$. In both cases we choose cut-off levels so that $P(Y_i = 1) = \pi$, $\forall i$. Note that the correlation matrix R is identical in both models (it is given by a equicorrelation matrix with off-diagonal element ρ). However, the copula of $\mathbf{X}$ differs, and we expect more joint defaults in the t model, due to the tail dependence of the t copula.

We define 3 groups of decreasing credit quality, labelled A, B and C.[1] In Group A we set $\pi = 0.06\%$ and $\rho = 2.58\%$; in Group B we set $\pi = 0.50\%$ and $\rho = 3.80\%$; in Group C we set $\pi = 7.50\%$ and $\rho = 9.21\%$. For each group we vary portfolio size m and, most importantly, the degree of freedom parameter ν. To perform the simulation we use the equivalent Bernoulli mixture model representations with default probabilities given in (18). For m large this is more efficient than simulating the latent variables directly, as m independent Bernoulli variates are easier to simulate than an equal amount of independent normal variates. In order to represent the tail of the number of defaults M we estimate empirically the 95% and 99% quantiles $\widehat{q_{0.95}}(M)$ and $\widehat{q_{0.99}}(M)$ and tabulate them in Table 1.

Clearly ν has a massive influence on the high quantiles, particularly for groups of poorer credit quality (B and C). This indicates that any attempt to calibrate latent variable models, which is solely based on marginal default probabilities and assumptions about latent variable correlations, is not advisable, as the remaining model risk related to the choice of an appropriate copula can be substantial.

The impact of the form of the mixing distribution in Bernoulli mixtures Next we look at the model risk related to different specifications of the mixing-distribution in various exchangeable Bernoulli mixture models, assuming that default probability π and

[1] These groups do not correspond exactly to the A, B and C rating categories used by any of the well-known rating agencies, but they are nonetheless realistic values for Gaussian latent variable models for real obligors and were chosen after discussions with a Swiss bank.

m	Group	$\widehat{q_{0.95}}(M)$				$\widehat{q_{0.99}}(M)$			
		$\nu=\infty$	$\nu=50$	$\nu=10$	$\nu=4$	$\nu=\infty$	$\nu=50$	$\nu=10$	$\nu=4$
1000	A	2	3	3	0	3	6	13	12
1000	B	12	16	24	25	17	28	61	110
1000	C	163	173	209	261	222	241	306	396
10000	A	14	23	24	3	21	49	118	126
10000	B	109	153	239	250	157	261	589	1074
10000	C	1618	1723	2085	2587	2206	2400	3067	3916

Table 1: Results of Simulation study. Estimated 95th and 99th percentiles of the distribution of M in an exchangeable model. See text for the values of π and ρ corresponding to the 3 groups A, B and C. Note that the quantiles are approximately proportional to the size of the portfolio; this shows that the asymptotic result of Proposition 4.8 is useful even for relatively small portfolios.

default-correlation ρ_Y (or equivalently π and π_2) are known and fixed. According to Proposition 4.8 the tail of M is essentially determined by the tail of Q. In Figure 1 we plot the tail function of the probit-normal mixing model the logit-normal model and the beta-model on a logarithmic scale.[2] Inspection of Figure 1 shows that the distributions diverge only after the 99% quantile, the logit-normal mixing-distributions being the one with the heaviest tail. From a practical point of view this means that the particular parametric form of the mixing distribution in a Bernoulli-mixture model is of minor importance once π and ρ_Y have been fixed. Of course this does not mean that Bernoulli-mixtures are immune to model-risk; the tail of M is quite sensitive to π and in particular to ρ_Y, and these parameters are not easily estimated; see Section 5.2 for a useful approach.

Remark 5.1. Another approach to measuring the model-risk in exchangeable Bernoulli-mixtures with given π and ρ_Y is to look for asymptotically worst mixing distributions. Here a mixing distribution μ on $[0,1]$ with df G_μ is called asymptotically worse than a mixing distribution G_ν with df ν, if the tail of G_μ dominates the tail of G_ν, i.e. if there is some $\delta > 0$ such that $G_\mu(q) < G_\nu(q)$ for all $q \in [1-\delta, 1)$. Clearly, this definition is motivated by the observation that the tail of the mixing variable determines the tail of M.

Given two mixing distributions μ and ν it is immediate that μ is asymptotically worse than ν if $\mu(\{1\}) > \nu(\{1\})$. Hence we may look for an asymptotically worst mixing distribution, i.e. for a distribution μ^* maximizing $\mu(1)$ among all distributions on $[0,1]$ with first two moments equal to π and π_2 for given constants $\pi \in (0,1)$ and $\pi_2 \in [\pi^2, \pi]$. Interestingly, such a distribution can be determined explicitly; the asymptotically worst distribution is given by the two point-distribution

$$\mu_2^* = (1-p^*)\delta_{x^*} + p^*\delta_1\,, \text{ where } p^* := \frac{\pi_2 - \pi^2}{1 - 2\pi + \pi_2}\,,\ x^* := \frac{\pi - \pi_2}{1-\pi}\,, \tag{21}$$

and where δ_x denotes the Dirac-measure in the point x. For a proof we refer to Frey and McNeil (2001). It is easily seen that for realistic values of π and π_2 the mass $p^* = \mu^*(\{1\})$ is quite small; this confirms our earlier observation that by fixing π and π_2 the tail of the loss-distribution is to a large extent determined.

Systematic revovery risk. In standard portfolio risk models it is assumed that the loss given default is independent of the default event. However, one expects the loss given default to depend on the same risk factors as default-probabilities, so that we have systematic recovery risk. For example, the recovery rate on defaulted real estate loans depends on

[2] The tail-function of the Bernoulli mixture model corresponding to the t-model is not easily calculated; simulations, which we do not repeat, have shown that the behaviour of the t-model with fixed π and ρ_Y is simular to the behaviour of the other models.

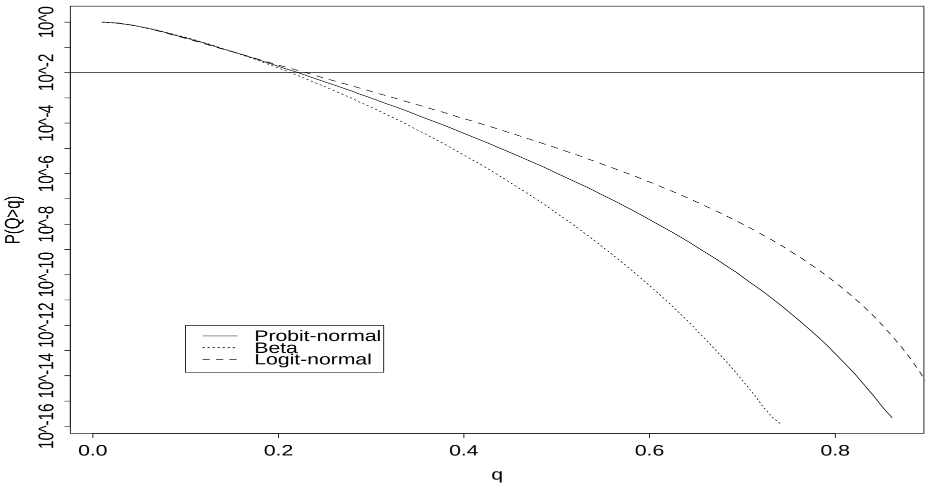

Figure 1: Tail of the mixing distribution G of Q in three different exchangeable Bernoulli mixture models: probit-normal; logit-normal; beta. In all cases we have $\pi = 0.075$ and $\pi_2 = 0.0076$. Horizontal line at 0.01 shows that models only really start to diverge at 99th percentile of mixing distribution.

the value of the real estate collateral, which is likely to be lower in situations where many other real estate projects have failed (Gordy 2001). The exact nature of this dependence is of course an empirical issue, which is beyond the scope of the present paper. Using the asymptotic results from Section 4.2, we may however assess qualitatively the impact of systematic recovery risk on quantiles of the portfolio loss distribution in certain one-factor Bernoulli mixture models.

Assume that the asymptotic conditional loss function $\overline{\ell}$ introduced in Assumption 4.4 is of the multiplicative form $\overline{\ell}(\psi) = \overline{\Delta}(\psi)\overline{Q}(\psi)$ for increasing functions $\overline{\Delta}$ and $\overline{Q}$; cf Remark 4.7. By Proposition 4.8 we get

$$q_\alpha(L^{(m)}) \approx m\overline{\ell}(q_\alpha(\Psi)) = m\overline{Q}(q_\alpha(\Psi))\overline{\Delta}(q_\alpha(\Psi)).$$

Consider alternatively a model with identical conditional default probabilities but with asymptotic conditional loss of the form $\widetilde{\ell}(\Psi) = \widetilde{\Delta}\,\overline{Q}(\Psi)$ for a constant $\widetilde{\Delta}$, which is independent of Ψ. In order to compare the two models we assume that the expected asymptotic losses are equal in both model, i.e. that $E\left(\widetilde{\ell}(\Psi)\right) = E\left(\overline{\ell}(\Psi)\right)$. This yields for α close to one

$$\widetilde{\Delta} = \frac{E\left(\overline{\Delta}(\Psi)\overline{Q}(\Psi)\right)}{E\left(\overline{Q}(\Psi)\right)} \overset{\alpha\approx 1}{\leq} \frac{E\left(\overline{\Delta}(q_\alpha(\Psi))\overline{Q}(\Psi)\right)}{E\left(\overline{Q}(\Psi)\right)} = \overline{\Delta}(q_\alpha(\Psi))\,.$$

Moreover, we have $q_\alpha\big(\widetilde{L}^m\big) \approx m\,\widetilde{\Delta}\,\overline{Q}(q_\alpha(\Psi))$ and hence $q_\alpha\left(L^{(m)}\right)/q_\alpha\big(\widetilde{L}^m\big) \approx \overline{\Delta}(q_\alpha(\Psi))/\widetilde{\Delta}$. As this is bigger than one for α close to one, we see that systematic recovery risk does indeed increase high quantiles of the loss distribution, at least in the case considered here.

5.2 Calibration of one-factor Bernoulli mixtures

Suppose we wish to calibrate a suitable one-factor Bernoulli mixture model for a group of obligors whose default risks are summarised by a simple credit rating system. We can calibrate such a model for a one-year time horizon by statistically fitting it to historical data on numbers of yearly defaults in each rating category, subject to the important caveat that such data are available. The relevant data would depend on whether we used an internal rating system or adopted a rating agency system, such as that of Standard and Poor's or

Moodys. An internal system would require us to keep our own database of annual numbers of defaults in each rating category; otherwise we could use data provided by the external rating agency. We will give a concrete example of the method at the end of this section using data collected in Standard and Poor's (2001).

Suppose we have historical data for the years $j = 1, \ldots, n$ for k different rating classes indexed by $r = 1, \ldots, k$. In year j the cohort consists of $m_{j,r}$ obligors in rating class r, of which $M_{j,r}$ default in the course of the year. (We treat $m_{j,r}$ as a fixed constant and $M_{j,r}$ as a rv.) We assume that in year j the conditional distribution of $\mathbf{M}_j = (M_{j,1}, \ldots, M_{j,k})'$ is of the form (14) given a standard normally distributed factor variable Ψ_j. The unconditional distribution of $\mathbf{M}_j$ is given by

$$P(\mathbf{M}_j = \mathbf{l}_j) = \int_{-\infty}^{\infty} \prod_{r=1}^{k} \binom{m_{j,r}}{l_{j,r}} \left(h(\mu_r + \sigma_r z)\right)^{l_{j,r}} \left(1 - h(\mu_r + \sigma_r z)\right)^{m_{j,r} - l_{j,r}} \phi(z) dz, \qquad (22)$$

where $\mathbf{l}_j = (l_{j,1}, \ldots, l_{j,k})'$ and where ϕ denotes the the standard normal density.

To complete the specification we assume that the factor rvs $\Psi_1, \ldots, \Psi_n$ for each of the different years are iid standard normal. Note that Ψ_j represents the *change* in the value of the factor (eg. a macroeconomic variable) over year j, so that the independence assumption is consistent with a model where the *level* of the factor follows a simple autoregressive structure, which attempts to capture economic cycles. We also assume that for $j_1 \neq j_2$ $\mathbf{M}_{j_1}$ and $\mathbf{M}_{j_2}$ are conditionally independent given Ψ_{j_1} and Ψ_{j_2}. Thus the joint distribution of $\mathbf{M}_1, \ldots, \mathbf{M}_n$ is the product of their marginal distributions as given in (22). This allows us to write down the log-likelihood for the unknown parameters of this model, which are $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ and $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_k)'$, in terms of the data $\mathbf{l}_1, \ldots, \mathbf{l}_n$, which are realisations of $\mathbf{M}_1, \ldots, \mathbf{M}_n$. We have

$$L(\boldsymbol{\mu}, \boldsymbol{\sigma}; \mathbf{l}_1, \ldots, \mathbf{l}_n) = \sum_{j=1}^{n} \sum_{r=1}^{k} \log \binom{m_{j,r}}{l_{j,r}} + \sum_{j=1}^{n} \log I_j$$

where $I_j = \int_{-\infty}^{\infty} \prod_{r=1}^{k} \left(h(\mu_r + \sigma_r z)\right)^{l_{j,r}} \left(1 - h(\mu_r + \sigma_r z)\right)^{m_{j,r} - l_{j,r}} \phi(z) dz$.

Maximisation of the log-likelihood with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ requires n numerical integrations for every point at which the log-likelihood is evaluated. To avoid numerical problems we have found it useful to make the substitution $q = \Phi(z)$ and to rewrite and evaluate I_j as

$$I_j = \int_0^1 \exp\left(\sum_{r=1}^{k} l_{j,k} \log\left(h(\mu_r + \sigma_r \Phi^{-1}(q))\right) + (m_{j,r} - l_{j,r}) \log\left(1 - h(\mu_r + \sigma_r \Phi^{-1}(q))\right) \right) dq.$$

Having fitted such a model and obtained estimates $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\sigma}}$, we can easily infer estimates of default probabilities as well as within-group and between-group default correlations for each of the groups. In fact fitting the one-factor Bernoulli model can be viewed as a suitable estimation method for obtaining values for these important first and second order moments.

Let $\pi^{(r)}$ denote the default probability for obligors in group r and let $\widehat{Q}_r(\Psi) = h(\widehat{\mu}_r + \widehat{\sigma}_r \Psi)$ for a generic normally distributed factor rv Ψ; we have $\widehat{\pi}^{(r)} = \int_{-\infty}^{\infty} h(\widehat{\mu}_r + \widehat{\sigma}_r z)\phi(z) dz$. Defining

$$\widehat{\pi}_2^{(r,s)} := E\left(\widehat{Q}_r(\Psi)\widehat{Q}_s(\Psi)\right) = \int_0^1 \left(h(\widehat{\mu}_r + \widehat{\sigma}_r \Phi^{-1}(q)) h(\widehat{\mu}_s + \widehat{\sigma}_s \Phi^{-1}(q))\right) dq,$$

for $1 \leq r, s \leq k$, the $k \times k$ matrix of estimated within-group and between-group default correlations has (r, s)-element given by

$$\widehat{\rho}_Y^{(r,s)} = \frac{\widehat{\pi}_2^{(r,s)} - \widehat{\pi}^{(r)} \widehat{\pi}^{(s)}}{\sqrt{(\widehat{\pi}^{(r)} - \widehat{\pi}^{(r)2})(\widehat{\pi}^{(s)} - \widehat{\pi}^{(s)2})}}.$$

The diagonal elements $\widehat{\rho}_Y^{(r,r)}$ are the estimated within-group default correlations, i.e. the correlation of default indicators for any two obligors in group r.

Group r	μ_r	(s.e.)	σ_r	(s.e.)	$\pi^{(r)}$	$\rho_Y^{(r,r)}$
A	-3.40	0.14	0.189	0.17	0.0004	0.0002
BBB	-2.90	0.09	0.205	0.10	0.0022	0.0010
BB	-2.41	0.08	0.252	0.07	0.0098	0.0048
B	-1.69	0.06	0.239	0.05	0.0503	0.0130
C	-0.84	0.08	0.262	0.07	0.2066	0.0327

Table 2: Maximum likelihood parameter estimates and standard errors for a one-factor Bernoulli mixture model fitted to historical Standard and Poor's one-year default data, together with the estimated default probabilities $\widehat{\pi}^{(r)}$ and estimated within-group default correlations $\widehat{\rho}_Y^{(r,r)}$ that these imply.

An Example with Standard and Poor's Data. In Standard and Poor's (2001) (see Table 13 on pages 18-21) one-year default rates for groups of obligors formed into cohorts (described as static pools) in the years 1981-2000 can be found. From this information it is possible to infer the actual numbers of defaulting obligors. Standard and Poor's use the ratings AAA, AA, A, BBB, BB, B, CCC, but because the one-year default rates for AAA and AA-rates obligors are largely zero, we concentrate on the rating categories A to CCC where defaults over the one-year horizon are observed. Thus we work with $n = 20$ years of data and $k = 5$ rating classes.

The parameter estimates obtained by maximum likelihood in the case of the probit link function are given in Table 2, together with the estimated default probabilities $\widehat{\pi}^{(r)}$ and estimated within-group default correlations $\widehat{\rho}_Y^{(r,r)}$ that these imply. In Table 3 the full matrix of estimated within-group and between-group default correlations is shown. These values may prove a useful reference for researchers who seek plausible values for default correlations in other studies of portfolio credit risk.

$\rho_Y^{(r,s)}$	A	BBB	BB	B	C
A	0.00022	0.00047	0.00103	0.00166	0.00256
BBB	0.00047	0.00103	0.00223	0.00361	0.00564
BB	0.00103	0.00223	0.00484	0.00791	0.01226
B	0.00166	0.00361	0.00791	0.01303	0.02048
C	0.00256	0.00564	0.01226	0.02048	0.03270

Table 3: Implied estimates of within-group and between-group default correlations based on maximum-likelihood fitting of a one-factor Bernoulli mixture model to Standard and Poor's historical one-year default data.

The use of the maximum likelihood method now permits standard likelihood-based inferential procedures to be applied. For example, we might investigate whether our 10-parameter model could be replaced by a more parsimonious model with less parameters. It may be that certain rating classes could be amalgamated and certain parameter values set to be equal. This amounts to considering simplified models nested within our basic model and such hypotheses can be addressed using likelihood ratio tests.

5.3 Calibration of normal variance mixtures

We now turn our attention to the calibration of models for more heterogeneous groups of obligors constructed using the latent variable philosophy underlying KMV and CreditMetrics. We assume the factor structure of the latent variables $\mathbf{X}$ has dimension greater than one and the dispersion matrix Σ of the latent variables has rich structure. This renders the approach of trying to statistically fit an equivalent mixture model to historical default data practically impossible, since relevant data will be scarce.

Under these circumstances other approaches to model calibration are used. Default probabilities are either inferred using internal or external ratings or, via variants of the Merton (1974) model from equity values. The parameters describing the factor model are chosen either by an ad-hoc consideration of which factors influence asset returns and to what extent, or possibly by a more formal regression analysis of asset returns (or a proxy like equity returns) against economic factors. In the spirit of this approach we will assume then that individual default probabilities and the factor structure of $\mathbf{X}$ as summarised by the matrix Σ are given.[3] For a Gaussian latent variable model this completes the calibration but, as we have seen in Section 5.1, the remaining model risk may be substantial.

If we extend the Gaussian model to some parametric family of normal variance mixture models, we introduce new parameters for the mixing variable W and hence for the amount of tail dependence of the latent variables. In this section we investigate how these additional parameters could be statistically estimated from historical default data, under the assumption that all other parameters are given. In our opinion this approach could be useful to reduce the model risk. We obtain a model which combines a rational calibration based on consideration of relevant economic factors affecting default as well as an adjustment to improve the fit of the model to observed historical defaults.

To accommodate default observation over successive years, indexed again by $j = 1, \dots, n$, we introduce a multiperiod version of our variance mixture model. Denote by m_j, $1 \le j \le n$ the number of obligors in the sample of observed firms in year j. In any given year the default indicator vector $\mathbf{Y_j} = (Y_{j,1}, \dots, Y_{j,m_j})'$ is induced by a latent variable model $(X_{j,i}, D_0^{j,i})$, $1 \le i \le m_j$, where the latent variable vector $\mathbf{X}_j$ follows a normal variance mixture model with factor structure. We assume the following.

Assumption 5.2.

(i) $\mathbf{X}_j = g(W)\mathbf{Z}_j$, where $\mathbf{Z}_j$ follows the p factor model $Z_{j,i} = (A_j\boldsymbol{\Theta}_j)_i + \sigma_{j,i}\varepsilon_{j,i}$, $\boldsymbol{\Theta_j} \sim N_p(\mathbf{0}, \Omega_j)$ are systematic effects, $\boldsymbol{\varepsilon_j} \sim N_{m_j}(\mathbf{0}, I)$ are idiosyncratic effects, and $\boldsymbol{\Theta_j}$ and $\boldsymbol{\varepsilon_j}$ are independent. The coefficients of the matrices A_j and Ω_j as well as the parameters $\sigma_{j,i}$, which determine the factor structure of $\mathbf{X}_j$, are known. The thresholds $D_0^{j,i}$ are chosen such that the unconditional default probabilities $P(Y_{j,i} = 1)$ are equal to $\overline{p}_{j,i}$ for some array $\overline{p}_{j,i}$ of estimates for the default probability of firm i in year j (made at the beginning of year j).

(ii) $(W_j)_{1\le j\le n}$ is an iid sequence of rv's with distribution depending on an unknown parameter vector $\boldsymbol{\nu}$ from some set $\mathcal{M}^{\boldsymbol{\nu}}$; the df of a generic W is denoted by $G_{\boldsymbol{\nu}}(w)$.

(iii) $(\boldsymbol{\varepsilon}_j, \boldsymbol{\Theta}_j)_{1\le j\le n}$ is an iid sequence of random vectors; moreover, the sequence $(W_j)_{1\le j\le n}$ is independent of the sequence $(\boldsymbol{\varepsilon}_j, \boldsymbol{\Theta}_j)_{1\le j\le n}$.

Under Assumption 5.2 the conditional default probabilities in year j are given by

$$P(Y_{j,i} = 1 \mid \boldsymbol{\Theta_j} = \boldsymbol{\theta}, W_j = w) = Q_{j,i}(\boldsymbol{\theta}, w) := \Phi\left(\frac{D_0^{j,i} - g(w)(A_j\boldsymbol{\theta})_i}{g(w)\sigma_{j,i}}\right). \tag{23}$$

Maximum likelihood estimation. To implement the ML-method we require the joint probability function of all our default observations. The default indicators $\mathbf{Y}_j$, $1 \le j \le n$ form a sequence of independent random vectors. Denote by $\mathbf{y}_j = (y_{j,1}, \dots, y_{j,m_j})'$ the default-observations in year j. Let $A_j = \{1 \le i \le m_j : y_{j,i} = 1\}$ be the identities of the firms which have defaulted in year j, and $A_j^c := \{1, \dots, m_j\} - A_j$ the identities of the surviving firms. Using the conditional independence of the default indicators we obtain

$$P(\mathbf{Y}_j = \mathbf{y}_j \mid W_j, \boldsymbol{\Theta}_j) = \prod_{i\in A_j} P(Y_{j,i} = 1 \mid \boldsymbol{\Theta_j}, W_j) \prod_{i\in A_j^c} (1 - P(Y_{j,i} = 1 \mid \boldsymbol{\Theta_j}, W_j)).$$

[3]Note that this is exactly the information which is provided by the KMV-amodel to subscribers of the service.

The unconditional default probability function of the observations depends on the unknown parameters of the model $\boldsymbol{\nu}$ and is given by

$$P(\mathbf{Y}_j = \mathbf{y}_j\,;\boldsymbol{\nu}) = \int_{\mathbb{R}} \int_{\mathbb{R}^p} P(\mathbf{Y}_j = \mathbf{y}_j \mid \boldsymbol{\Theta}_j = \boldsymbol{\theta}, W_j = w) dF_{\mathbf{0},\Omega_j}(\boldsymbol{\theta}) dG_{\boldsymbol{\nu}}(w)\,,$$

where $F_{\mathbf{0},\Omega_j}(\boldsymbol{\theta})$ denotes the joint df of a $N_p(\mathbf{0},\Omega_j)$-distributed random vector.

Typically, the integral on the rhs has to be evaluated using MC-simulation. Under our independence assumptions the log-likelihood function for given observations $\mathbf{y}_1,\dots,\mathbf{y}_n$ equals $L(\boldsymbol{\nu};\mathbf{y}_1,\dots,\mathbf{y}_n) = \sum_{j=1}^n \ln P(\mathbf{Y}_j = \mathbf{y}_j\,;\boldsymbol{\nu})$. Following standard likelihood theory the ML-estimator for $\boldsymbol{\nu}$ is now given by

$$\widehat{\boldsymbol{\nu}}(\mathbf{y}_1,\dots,\mathbf{y}_n) = \arg\max\{L(\boldsymbol{\nu};\mathbf{y}_1,\dots,\mathbf{y}_n) \,:\, \boldsymbol{\nu} \in \mathcal{M}^{\boldsymbol{\nu}}\}\,. \tag{24}$$

Remark 5.3. 1) If the number of obligors is large, it is advisable to change default probabilities and factor structure slightly (for the purpose of estimating $\boldsymbol{\nu}$), so that the portfolio can be divided into a small number of homogeneous groups with equal unconditional default probability and equal factor structure. The basic default observation now consists of the number of firms in a group which actually defaulted in a given year j. The log-likelihood function and the MLE-estimator $\widehat{\boldsymbol{\nu}}$ can be computed as in the standard case. The main advantage of this approach is better numerical performance of the MC-simulation in the computation of the log-likelihood function.
2) To further enhance the performance of the MC-simulation variance-reduction methods can be used. We used a conditional importance sampling technique for our simulation study. Since the approach is useful for working with normal mixture models in general, we briefly sketch the idea. To compute the expectation $E(f(\mathbf{X}))$, where $\mathbf{X} = g(W)\mathbf{Z}$ follows a normal mixture model with $\mathbf{Z} \sim N(\mathbf{0},\Sigma)$ and $\Sigma = AA'$, we use the identity

$$\begin{aligned} E(f(\mathbf{X})) &= \int_{\mathbb{R}} \int_{\mathbb{R}^d} f(g(w)A\tilde{\mathbf{z}})\, dF_{\mathbf{0},I}(\tilde{\mathbf{z}})\; dG(w) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}^d} \frac{dF_{\mathbf{0},I}(\tilde{\mathbf{z}})}{dF_{\boldsymbol{\mu}(w),I}(\tilde{\mathbf{z}})} f(g(w)A\tilde{\mathbf{z}})\, dF_{\boldsymbol{\mu}(w),I}(\tilde{\mathbf{z}})\; dG(w)\,. \end{aligned}$$

The new mean $\boldsymbol{\mu}(w)$ of the relocated multivariate normal distribution with df $F_{\boldsymbol{\mu}(w),I}$, which may depend on the realization w of the mixing variable W, is chosen so that the second moment of the inner integral is reduced. This can be done using standard importance sampling techniques for multivariate normal rv's; in particular the density ratio $dF_{\mathbf{0},I}/dF_{\boldsymbol{\mu}(w),I}$ is easily computed.

A simulation study. To study the performance of the proposed estimator we ran a MC-simulation study with K trials indexed by $k = 1,\dots,K$. In each trial we simulated n years of default observations from a multivariate t model with known factor structure and applied MLE estimation to estimate the df parameter ν from these data, resulting in an estimate $\hat{\nu}_k$. Given $\hat{\nu}_k$ we computed the corresponding 99% quantile $q_{0.99}(M;\hat{\nu}_k)$ of the number of defaults. Since the tail of the loss distribution is the key object of interest in credit portfolio management, a density plot of the estimated values $q_{0.99}(M;\hat{\nu}_k)$, $k = 1,\dots,K$ permits a good assessment of the performance of the estimator.

We put $K = 100$ and $n = 20$, i.e. we tested the estimators assuming that we have 20 years of default observations. Our test portfolio consists of 400 obligors belonging to three homogeneous groups with different exposure to two systematic factors. We put the unknown parameter ν equal to 10; under our model assumptions we obtained $q_{0.99}(M;10) \approx 80$. In Figure 2 we graphed a density plot of $q_{0.99}(M;\hat{\nu}_k)$. We see from inspection that the MLE estimation method performs reasonably well in view of the small size of the simulated "sample" in each trial. In particular, given that in a normal model $q_{0.99}(M;\infty) \approx 58$ the

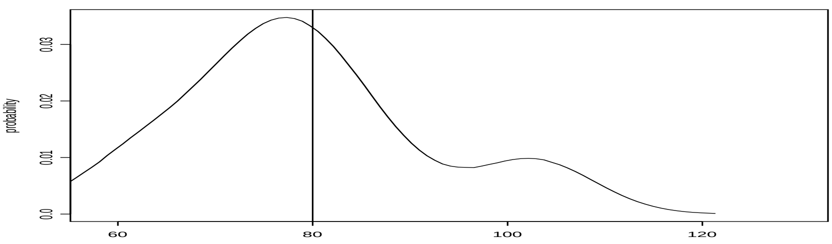

Figure 2: Density plot for the 99% quantile of M corresponding to an estimated value of the df parameter ν in a t-model with factor structure. The true value is approximately equal to 80. Details are given in Section 5.3.

estimator is generally capable to distinguish the t model with $\nu = 10$ from a normal model.[4] This suggests that in the context of a normal variance mixture models MLE estimation might be useful in reducing the model risk related to the choice of the copula.

As an alternative to MLE one could use moment-based estimators for $\boldsymbol{\nu}$. Theoretical moments can be computed via MC-simulation, using (16); empirical moments can either be estimated from historical data or using some ad hoc measure of portfolio concentration such as the "diversification score" put forward by Moody's (see for instance Davis and Lo (2001) for details on this measure). We tried this approach but found that with as little as $n = 20$ years of "observations" it performs significantly worse than MLE.

6 Conclusion

Ultimately, the goal of academic research on credit risk models is to help the practitioner in specifying a model, which is appropriate for his or her lending portfolio. We therefore conclude with a few recommendations on model choice, which summarize the more practical aspects of our research.

(Bernoulli) mixture models are handier than latent variable models when it comes to estimation and in particular simulation, at least for large portfolios. In case one prefers to work with a latent variable model, the model should have a conditional independence structure, so that it admits an equivalent representation as mixture model. As we have shown in Section 4.3, there is a wide range of latent variable models with this property.

One-factor regression models such as the model considered in Example 4.3 are reasonable models for portfolios with a relatively homogeneous exposure to a common set of risk factors. They are also useful parsimoneous models in a situation where we have only rather imprecise information on the risk factors affecting a portfolio, such that we have to solely on historical default information in estimating model parameters. As we have seen in Section 5.1, the precise form of the link-function (eg. logit or probit) is of lesser importance as long as the model is reasonably calibrated; hence a simple model such as a probit model should be used. For model calibration we recommend the MLE-approach proposed in Section 5.2.

[4] Theoretically $q_{0.99}(M; \infty)$ is a lower bound for the distribution of $q_{0.99}(M; \hat{\nu})$; the mass below 58 in Figure 2 is due to the kernel-estimator for the density.

For a portfolio, where obligors are exposed to several risk factors and where the exposure to different risk factors differs markedly accross obligors such as a portfolio of loans to larger corporations active in different industries or countries, we recommend a normal variance mixture model with factor structure. Calibration of the factor structure can be done by an analysis of asset returns or a proxy such as equity returns; the parameters of the mixing distribution can be estimated from historical default data using the approach developed in Section 5.3. This helps to reduce the large amount of model risk in Gaussian latent variable models, which have been calibrated solely to asset return data (see again Section 5.1).

References

Credit-Suisse-Financial-Products (1997): "CreditRisk$^+$ a Credit Risk Management Framework," Technical Document, available from `htpp://www.csfb.com/creditrisk`.

Crouhy, M., D. Galai, and R. Mark (2000): "A comparative analysis of current credit risk models," *Journal of Banking and Finance*, 24, 59–117.

Davis, M., and V. Lo (2001): "Infectious defaults," *Quantitative Finance*, 1, 382–387.

Duffie, D., and K. Singleton (1999): "Modeling Term Structure Models of Defaultable Bonds," *Review of Financial Studies*, 12, 687–720.

Eberlein, E., and U. Keller (1995): "Hyperbolic Distributions in Finance," *Bernoulli*, 1, 281–299.

Embrechts, P., A. McNeil, and D. Straumann (2001): "Correlation and dependency in risk management: properties and pitfalls," in *Risk Management: Value at Risk and Beyond*, ed. by M. Dempster, and H. Moffatt. Cambridge University Press.

Frey, R., and A. McNeil (2001): "Modelling dependent defaults," Preprint, ETH Zürich, available from `http://www.math.ethz.ch/~mcneil`.

Frey, R., and A. McNeil (2002): "VaR and Expected Shortfall in Portfolios of Dependent Credit Risks: Conceptual and Practical Insights," *Journal of Banking and Finance*, To appear.

Giesecke, K. (2001): "Structural modelling of defaults with incomplete information," preprint, Humboldt-Universität Berlin.

Gordy, M. (2000): "A comparative anatomy of credit risk models," *Journal of Banking and Finance*, 24, 119–149.

Gordy, M. (2001): "A Risk-Factor model foundation for ratings-based capital rules," working paper, Board of Governors of the Federal Reserve System.

Jarrow, R. and Yu, F. (2001): "Counterparty risk and the pricing of defaultable securities," *Journal of Finance*, 53, 2225–2243.

Joe, H. (1997): *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.

Kimberling, C. (1974): "A probabilistic interpretation of complete monotonicity," *Aequationes Mathematicae*, 10, 152–164.

KMV-Corporation (1997): "Modelling Default Risk," Technical Document, available from `http://www.kmv.com`.

Lando, D. (1998): "Cox processes and credit risky securities," *Review of Derivatives Research*, 2, 99–120.

Li, D. (2001): "On default correlation: a Copula function approach," *Journal of Fixed Income*, 9, 43–54.

Lindskog, F. (2000): "Modelling dependence with Copulas," RiskLab Report, ETH Zurich.

Marshall, A. and Olkien, I. (1988): "Families of multivariate distributions," *J. Am. Stat. Assoc.*, 83, 834–841.

Merton, R. (1974): "On the pricing of corporate debt: the risk structure of interest rates," *Journal of Finance*, 29, 449–470.

Nelsen, R. B. (1999): *An Introduction to Copulas*. Springer, New York.

Petrov, V. V. (1975): *Sums of Independent Random Variables*. Springer, Berlin.

RiskMetrics-Group (1997): "CreditMetrics – Technical Document," available from `http://www.riskmetrics.com/research`.

Schönbucher,P. (2002): "Taken to the limit: simple and not so simple loan loss distributions," working paper, department of statistics, faculty of economics, Universität Bonn.

Schönbucher, P., and D. Schubert (2001): "Copula-dependent default risk in intensity models," working paper, department of statistics, faculty of economics, Universität Bonn.

Schweizer, B., and A. Sklar (1983): *Probabilistic Metric Spaces*. North–Holland/Elsevier, New York.

Standard, and Poor's (2001): "Ratings Performance 2000: Default, Transition, Recovery, and Spreads," .

Wilson, T. (1997): "Portfolio Credit Risk I and II," *Risk*, 10(Sept and Oct).

A Copulas

In the following we present a brief introduction to copulas. For further reading see Embrechts, McNeil, and Straumann (2001), Joe (1997) and Nelsen (1999).

Definition A.1 (Copula). A copula is a multivariate distribution with standard uniform marginal distributions, or the df of such a distribution.

We use the notation $C(\mathbf{u}) = C(u_1, \dots, u_d)$ for the d-dimensional joint dfs which are copulas. C is a mapping of the form $C : [0,1]^d \to [0,1]$, i.e. a mapping of the unit hypercube into the unit interval. The following three properties characterise a copula C.

1. $C(u_1, \dots, u_d)$ is increasing in each component u_i.
2. $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$ for all $i \in \{1, \dots, d\}$, $u_i \in [0,1]$.
3. For all $(a_1, \dots, a_d), (b_1, \dots, b_d) \in [0,1]^d$ with $a_i \le b_i$ we have:

$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d} C(u_{1i_1}, \dots, u_{di_d}) \ge 0,$$

   where $u_{j1} = a_j$ and $u_{j2} = b_j$ for all $j \in \{1, \dots, d\}$.

Suppose the random vector $\mathbf{X} = (X_1, \dots, X_d)'$ has a joint distribution F with *continuous* marginal distributions $F_1, \dots, F_d$. If we apply the appropriate probability transform to each component we obtain a transformed vector $(F_1(X_1), \dots, F_d(X_d))$ whose df is by definition a copula, which we denote C. It follows that

$$\begin{aligned} F(x_1, \dots, x_n) &= P\left(F_1(X_1) \leq F_1(x_1), \dots, F_d(X_d) \leq F_d(x_d)\right) \\ &= C(F_1(x_1), \dots, F_d(x_d)), \end{aligned} \tag{25}$$

or alternatively $C(u_1, \dots, u_n) = F\left(F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d)\right)$, where $F_i^{\leftarrow}$ denotes the generalised inverse of the df F_i. Formula (25) shows how marginal distributions are *coupled together* by a copula to form the joint distribution and is the essence of Sklar's theorem.

Theorem A.2 (Sklar's Theorem). *Let F be a joint distribution function with margins $F_1, \dots, F_d$. Then there exists a copula $C : [0,1]^d \to [0,1]$ such that for all $x_1, \dots, x_d$ in $\overline{\mathbb{R}} = [-\infty, \infty]$ (25) holds; C is unique if $F_1, \dots, F_d$ are continuous. Conversely, if C is a copula and $F_1, \dots, F_d$ are distribution functions, then the function F given by (25) is a joint distribution function with margins $F_1, \dots, F_d$.*

For a proof we refer to Schweizer and Sklar (1983). If F is a joint df with marginals $F_1, \dots, F_d$ and (25) holds, we say that C a copula of F (or of a random vector $\mathbf{X} \sim F$).

A useful property of the copula of a distribution is its invariance under strictly increasing transformations of the marginals. Let $(X_1, \dots, X_d)$ be a vector of continuously distributed risks with copula C and let $T_1, \dots, T_d$ be strictly increasing functions. Then it is easily seen that $(T_1(X_1), \dots, T_d(X_d))$ also has copula C.

Random variables $X_1, \dots, X_d$ with continuous marginals are independent if and only if their copula is $C^{ind}(u_1, \dots, u_d) = \prod_{i=1}^d u_i$. Each of $X_1, \dots, X_d$ is almost surely a strictly increasing function of any of the others (a concept known as comonotonicity) if and only if their copula is $C^u(u_1, \dots, u_d) = \min\{u_1, \dots, u_d\}$. The copula of the d-dimensional Gaussian distribution takes the form

$$C_R^{\mathrm{Ga}}(\mathbf{u}) = \mathbf{\Phi}_R\left(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)\right), \tag{26}$$

where $\mathbf{\Phi}_R$ denotes the joint df of a standard d-dimensional normal random vector $\mathbf{X}$ with correlation matrix R, and Φ is the df of univariate standard normal. We simplify the notation to C_ρ^{Ga} in the case when all pairwise correlations of $\mathbf{X}$ are equal to ρ (in which case $\mathbf{X}$ is an exchangeable Gaussian vector).