Dependent Defaults 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

16th June 2003

Abstract

We analyse the mathematical structure of portfolio credit risk models with particular regard to the modelling of dependence between default events in these models. We explore the role of copulas in latent variable models (the approach that underlies KMV and CreditMetrics) and use non-Gaussian copulas to present extensions to standard industry models. We explore the role of the mixing distribution in Bernoulli mixture models (the approach underlying CreditRisk$^+$) and derive large portfolio approximations for the loss distribution. We show that all currently used latent variable models can be mapped into equivalent mixture models, which facilitates their simulation, statistical fitting and the study of their large portfolio properties. Finally we develop and test several approaches to model calibration based on the Bernoulli mixture representation; we find that maximum likelihood estimation of parametric mixture models generally outperforms simple moment estimation methods.

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 disproportionately 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 main goal of the present paper is to present a framework for analysing existing industry models, and various models proposed in the academic literature, with regard to the mechanisms they use to model dependence between defaults. These mechanisms are at

*We wish to thank Dirk Tasche, Mark Nyfeler, Filip Lindskog and Philipp Schönbucher for useful discussions.

least as important in determining the overall credit loss of a portfolio under the model, as are assumptions regarding default probabilities of the individual obligors in the portfolio. In previous papers (Frey, McNeil, and Nyfeler 2001, Frey and McNeil 2002) we have shown that portfolio credit models can be subject to considerable model risk. Small changes to the structure of the model or to the model parameters describing dependence can have a large impact on the resulting credit loss distribution and in particular its tail. This is worrying because credit models are extremely difficult to calibrate reliably, due to the relative scarcity of good data on credit losses.

In our analysis of mechanisms for dependent credit events we divide existing models into two classes: latent variable models such as KMV or CreditMetrics which essentially descend from the firm-value model of Merton (Merton 1974); Bernoulli mixture models such as CreditRisk$^+$ where default events have a conditional independence structure conditional on common economic factors. This division reflects the way these models are conventionally presented rather than any fundamental structural difference and the recognition that CreditMetrics (usually presented as a latent variable model) and CreditRisk$^+$ (a mixture model) can be mapped into each other goes back to Gordy (2000) and also Koyluoglu and Hickman (1998). In this paper we take this work one step further and develop a general result showing that in essentially all relevant cases for practical work the two model classes can be mapped into each other and thus reduced to a common framework. The more useful mapping direction is to rewrite latent variable models as Bernoulli mixture models, as there are a number of advantages to the latter presentation, which we discuss in this paper:

- Bernoulli mixture models are easy to simulate in Monte Carlo risk analyses. As a by-product of our analyses we obtain methods for simulating many of the models.
- Mixture models are more convenient for statistical fitting purposes. We show in this paper that maximum likelihood techniques can be applied.
- The large portfolio behaviour of Bernoulli mixtures can be understood in terms of the behaviour of the distribution of the common economic factors.

The recent literature contains a number of related papers beginning with the important paper of Gordy (2000). 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), and Schönbucher and Schubert (2001). The common theme of these papers is to construct models which reproduce realistic patterns of joint defaults. The last paper makes 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.

Our paper is organised as follows. In Section 2 we provide some general notation for describing the default models of this paper. We study latent variable models in Section 3; we show that existing industry models are essentially structurally similar and we use copulas to suggest how the class of latent variable models may be extended to get new dependence structures between defaults. Mixture models are considered in Section 4; here we show how asymptotic calculations for large portfolios can be made in the Bernoulli mixture framework and we give a general result for mapping between the two model classes. In Section 5 we discuss the statistical calibration of Bernoulli mixture models and Section 6 contains practical conclusions for practitioners. A short introduction to copula theory is included in Appendix A and the proofs of the propositions and lemmas appearing in the paper are found in Appendix B.

2 Models for loan portfolios

The division of credit models into latent variable and mixture models corresponds to usage of these terms in the statistics literature; see for example Joe (1997). In latent variable models default occurs if a random variable X (termed a 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 (Kealhofer and Bohn 2001, Crosbie and Bohn 2002) 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).

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 $1 \le i \le m$, let the random variable 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 credit-worthiness. 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 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 $0 \le \Delta_i \le 1$ 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 $\widetilde{\mathbf{S}}$ (or $\mathbf{Y}$ and $\widetilde{\mathbf{Y}}$) *equivalent* if $\mathbf{S} \stackrel{\mathrm{d}}{=} \widetilde{\mathbf{S}}$ (or $\mathbf{Y} \stackrel{\mathrm{d}}{=} \widetilde{\mathbf{Y}}$), 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 said to be 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 $(1, \dots, m)$. Note that this homogeneity only applies to the phenomenon of default and we might still have quite heterogeneous exposures and losses given default; even with heterogeneous exposures exchangeability remains useful as it simplifies specification and calibration of the model for defaults. Exchangeability 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.

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

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

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

In particular, the default correlation ρ_Y (i.e. 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 $d_1^i<\cdots<d_n^i$ be a sequence of *cut-off* levels. Put $d_0^i=-\infty$, $d_{n+1}^i=\infty$ and set

$$S_i=j \iff d_j^i<X_i\le d_{j+1}^i \quad j\in\{0,\dots,n\},\ i\in\{1,\dots,m\}.$$

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_1^i are often interpreted as the values of assets and liabilities respectively 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 $\overline{p}_i=F_i(d_1^i)$.

We now give a simple criterion for equivalence of two latent variable models in terms of the marginal distributions of the state vector $\mathbf{S}$ and the *copula* of $\mathbf{X}$; while straightforward from a mathematical viewpoint this result suggests a new way of looking at the structure of latent variable models and 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).

Lemma 3.2. *Let* $\left(X_i,(d_j^i)_{01\le j\le n}\right)_{1\le i\le m}$ *and* $\left(\widetilde{X}_i,(\widetilde{d}_j^i)_{1\le j\le n}\right)_{1\le i\le m}$ *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.*

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

(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 $(\bar{p}_i)_{1\le i\le m}$ 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.

We now give some examples of industry credit models which are all based implicitly on the Gaussian copula, the unique copula describing the dependence structure of the multivariate normal distribution. See Appendix A for a mathematical definition of this copula.

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_1^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 Kealhofer and Bohn (2001) and Crosbie and Bohn (2002). 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

$$X_i = \sum_{j=1}^{p} a_{i,j}\Theta_j + \sigma_i\varepsilon_i + \mu_i, \quad i = 1,\dots,d\,, \tag{3}$$

for some $p < m$, a p-dimensional random vector $\mathbf{\Theta} \sim N_p(\mathbf{0},\Omega)$ and independent standard normally distributed random variables $\varepsilon_1 \dots, \varepsilon_m$, which are also independent of $\mathbf{\Theta}$. This implies that Σ is of the form $\Sigma = A\Omega A' + \operatorname{diag}(\sigma_1^2,\dots,\sigma_m^2)$. 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. 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 $(\bar{p}_i)_{1\le i\le m}$ 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, which is set up in continuous time, 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. $F_i(t) = 1 - \exp(-\lambda_i t)$. Company i has defaulted by time T if and only if $X_i \le T$, so that $\bar{p}_i = F_i(T)$ and $(X_i, T)_{1\le i\le m}$ describes the latent variable model for $\mathbf{Y}$. 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 (25) 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. Dynamic properties of this model are studied in Schönbucher and Schubert (2001).

While most latent variable models popular in industry are based on the Gaussian copula, there is no reason why we have to assume a Gaussian copula. Alternative copulas can lead to very different credit loss distributions, and this may be understood by considering a subgroup of k companies $\{i_1, \ldots, i_k\} \subset \{1, \ldots, m\}$, with individual default probabilities $\overline{p}_{i_1}, \ldots, \overline{p}_{i_k}$ and observing that

$$P(Y_{i_1} = 1, \ldots, Y_{i_k} = 1) = P\left(X_{i_1} \leq d_1^{i_1}, \ldots, X_{i_k} \leq d_1^{i_k}\right) = C_{i_1,\ldots,i_k}\left(\overline{p}_{i_1}, \ldots, \overline{p}_{i_k}\right), \quad (4)$$

where $C_{i_1,\ldots,i_k}$ denotes the corresponding k-dimensional margin of C. It is obvious from (4) that the copula crucially determines joint default probabilities of groups of obligors and thus the tendency of the model to produce 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 across obligors such as a portfolio of loans to companies from different industries or countries. The following class of models preserves this feature of the KMV/CreditMetrics-type models and adds more flexibility.

Example 3.5 (Normal mean-variance mixtures). In this class we start with an m-dimensional multivariate normal vector $\mathbf{Z} \sim N_m(\mathbf{0}, \Sigma)$ and some random variable W, which is independent of Z. The latent variable vector $\mathbf{X}$ is assumed to have components of the form

$$X_i = \mu_i(W) + g(W)Z_i, \quad 1 \leq i \leq m, \quad (5)$$

for functions $\mu_i: \mathbb{R} \to \mathbb{R}$ and $g_i: \mathbb{R} \to (0, \infty)$. In the special case where μ_i is a constant not depending on W the distribution is called a normal variance mixture.

An example of a normal variance mixture is the multivariate t distribution with mean $\boldsymbol{\mu}$ and degrees of freedom ν, denoted by $t_m(\nu, \boldsymbol{\mu}, \Sigma)$. This is obtained from (5) by setting $\mu_i(W) = \mu_i$ for all i and $g(w) = \nu^{1/2}w^{-1/2}$, and then taking W to have a chi-squared random variable with ν degrees of freedom. This gives a distribution with t-distributed univariate marginals and covariance matrix $\frac{\nu}{\nu-2}\Sigma$. An example of a more general mean-variance mixture is the generalised hyperbolic distribution. Here we assume that the mixing variable W in (5) follows a so-called generalised inverse Gaussian distribution and take $\mu_i(W) = \beta_i W^2$ for constants β_i and $g(W) = W$. The generalised 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

$$X_i \leq d_1^i \iff Z_i \leq d_1^i h_1(W) - h_{i,2}(W) =: \widetilde{D}^i, \quad (6)$$

where $h_1(w) = 1/g(w)$ and $h_{i,2}(w) = \mu_i(w)/g(w)$. A possible economic interpretation of the model (5) is therefore to consider Z_i as asset value of company i and d_1^i as an a priori estimate of the corresponding default threshold. The actual default threshold is *stochastic* and is represented by $\widetilde{D}^i$, which is obtained by applying a multiplicative and an additive shock to the estimate d_1^i. If we interpret this shock as a stylised 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 random variable W. A similar idea underlies the model of Giesecke (2001).

Normal variance mixtures, such as the multivariate t, provide the most tractable examples and admit a similar calibration approach to models based on the Gaussian copula. In this class of models the correlation matrix of $\mathbf{X}$ (when defined) and $\mathbf{Z}$ coincide. Moreover, if $\boldsymbol{Z}$ follows the linear factor model (3), then $\mathbf{X}$ inherits the linear factor structure from $\mathbf{Z}$. Note however, that the "systematic factors" $g(W)\boldsymbol{\Theta}$ and the "idiosyncratic factors" $g(W)\varepsilon_i$, $1 \leq i \leq m$, are no longer independent but merely uncorrelated. A latent variable model based on the t copula (which we denote $C^t_{\nu,R}$) can be thought of as containing the standard

KMV/CreditMetrics model based on the Gauss copula C_R^{Ga} as a limiting case as $\nu \to \infty$. However the additional parameter ν adds a great deal of flexibility to the model, and it has been shown in Frey, McNeil, and Nyfeler (2001) that when the correlation matrix R of the latent variables is fixed, and even when ν is quite large, a model based on the t copula tends to give many more joint defaults than a model based on the Gaussian copula. This can be explained by the tail dependence of the t copula (see Embrechts, McNeil, and Straumann (2001) for a definition) which causes the t copula to generate more joint extreme values in the latent variables.

Alternatively we could use parametric copulas in closed-form to model the distribution of latent variables. An example is provided by the class of so-called Archimedean copulas.

Example 3.6 (Archimedean copulas). An Archimedean copula is the distribution function of an *exchangeable* uniform random vector and has the form

$$C(u_1, \ldots, u_m) = \phi^{-1}\left(\phi(u_1) + \cdots + \phi(u_m)\right), \tag{7}$$

where $\phi : [0,1] \mapsto [0,\infty]$ is a continuous, strictly decreasing function, known as the copula *generator* which satisfies $\phi(0) = \infty$ and $\phi(1) = 0$ and ϕ^{-1} is its inverse. In order that (7) defines a proper distribution for any portfolio size m the generator inverse must have the property of complete monotonicity (defined by $(-1)^m \frac{d^m}{dt^m}\phi^{-1}(t) \geq 0, \quad m \in \mathbb{N}$). There are many possibilities for generating Archimedean copulas (Nelsen 1999) and in this paper we will use as an example Clayton's copula which has generator $\phi_\theta(t) = t^{-\theta} - 1$, where $\theta > 0$. This gives the copula $C_\theta^{\mathrm{Cl}}(u_1, \ldots, u_m) = (u_1^{-\theta} + \ldots + u_m^{-\theta} + 1 - m)^{-1/\theta}$. Archimedean copulas suffer from the deficiency that they are not rich in parameters and can only model exchangeable dependence and not a fully flexible dependence structure for the latent variables. Nonetheless they yield useful parsimonious models for relatively small homogeneous portfolios, which are easy to calibrate and simulate as we discuss in more detail in Section 4.3.

Suppose $\mathbf{X}$ is a random vector with an Archimedean copula so that $\left(X_i, (d_j^i)_{1\leq j\leq n}\right)$, $1 \leq i \leq m$, specifies a latent variable model with individual default probabilities $F_i(d_1^i)$, where F_i denote the ith margin of $\mathbf{X}$. As a concrete example consider the Clayton copula and assume a homogeneous situation where all of these default probabilities are identical to π. Using the notation in (1) and relation (4), we can calculate that the probability that an arbitrarily selected group of k obligors from a portfolio of m such obligors defaults over the time horizon is given by $\pi_k = (k\pi^{-\theta} - k + 1)^{-1/\theta}$. Essentially the dependent default mechanism of the homogeneous group is now determined by this equation and the parameters π and θ. We study this Clayton copula model further in Examples 4.12 and 4.14.

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 $Q_i : \mathbb{R}^p \to [0,1]$, $1 \leq i \leq m$, 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

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

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

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

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

Here $\boldsymbol{\Psi} = (\Psi_1, \ldots, \Psi_p)'$ 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 $\boldsymbol{\Psi}$, the default of counterparty i occurs independently of other counterparties with a Poisson intensity given by $\Lambda_i(\boldsymbol{\Psi}) = \mathbf{w}_i'\boldsymbol{\Psi}$. 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 $\boldsymbol{\Psi}$ that a counterparty defaults over the time period of interest (whether once or more than once) is given by

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

so that we obtain the Bernoulli mixture model in (9). The Poisson formulation of CreditRisk$^+$ (together with the positivity of the factor weights) leads to 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 random variable 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.

4.1.1 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 random variable $Q := Q_1(\Psi)$ and to denote the distribution function 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,

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

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 (2) 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 random variable 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.

Example 4.3 (Beta, probit- and logit-normal mixtures). 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 distribution function. 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 and is inferred from (18).
- 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 a and b or μ and σ of the mixing distribution and higher order joint default probabilities are automatic.

4.1.2 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 tractable 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.

Clearly 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 4.4 (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

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

where $\boldsymbol{l} = (l_1, \ldots, l_k)'$. 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 4.7.

4.2 Loss distributions for large portfolios in Bernoulli mixture models

We now provide some asymptotic results for large portfolios in Bernoulli mixture models. Our results can be used for an approximate evaluation of the credit loss distribution in a large portfolio. Moreover, they will be useful in identifying the crucial parts of a Bernoulli mixture model. In particular, we will see that in one-factor models the tail of the loss distribution is essentially determined by the tail of the mixing distribution with direct consequences for the analysis of model risk in mixture models and for the setting of capital adequacy rules for loan books.

In this section we are interest in asymptotic properties of the loss given default so that we have to consider exposures and loss given default. Let $(e_i)_{i\in\mathbb{N}}$ be an infinite sequence of positive deterministic exposures, $(Y_i)_{i\in\mathbb{N}}$ be the corresponding sequence of default indicators and $(\Delta_i)_{i\in\mathbb{N}}$ a sequence of random variables with values in $(0, 1]$ representing percentages losses given that default occurs. In this setting the portfolio 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 on our model.

A1) There is a p-dimensional random vector $\boldsymbol{\Psi}$ and functions ℓ_i: $\mathrm{supp}(\boldsymbol{\Psi}) \to [0, 1]$ such that conditional on $\boldsymbol{\Psi}$ the $(L_i)_{i\in\mathbb{N}}$ form a sequence of independent random variables with mean $\ell_i(\boldsymbol{\Psi}) = E(L_i \mid \boldsymbol{\Psi})$.

In this assumption we extend the conditional independence structure from the default indicators to the losses. Note that in contrast to many standard models we do not assume that losses given default Δ_i and default indicators are independent.

A2) There is a function $\overline{\ell} : \text{supp}(\boldsymbol{\Psi}) \to \mathbb{R}^+$such that

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

for all $\boldsymbol{\psi} \in \text{supp}(\boldsymbol{\Psi})$, where $\text{supp}(\boldsymbol{\Psi})$ denotes the support of the distribution of $\boldsymbol{\Psi}$. We refer to $\overline{\ell}(\boldsymbol{\psi})$ as the asymptotic conditional loss function.

Assumption A2 implies that we preserve the essential composition of the portfolio as we allow it to grow; see for instance Example 4.7.

Our last assumption prevents exposures from growing systematically with portfolio size.

A3) There is some $C < \infty$ such that $\sum_{i=1}^{m} (e_i/i)^2 < C$ for all m.

The next result shows that under these assumptions the average portfolio loss is essentially determined by the realisation $\boldsymbol{\psi}$ of the economic factor variable $\boldsymbol{\Psi}$. A related result has independently been obtained by Gordy (2001).

Proposition 4.5. *Consider a sequence $L^{(m)} = \sum_{i=1}^{m} L_i$ satisfying Assumptions A1, A2 and A3 above. Denote by $P(\cdot \mid \boldsymbol{\Psi} = \boldsymbol{\psi})$ the conditional distribution of the sequence $(L_i)_{i\in\mathbb{N}}$ given $\boldsymbol{\Psi} = \boldsymbol{\psi}$. Then*

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

Proposition 4.5 obviously applies to the number of defaults $M^{(m)} = \sum_{i=1}^{m} Y_i$, if we put $\Delta_i = e_i \equiv 1$. For a given sequence $(Y_i)_{i\in\mathbb{N}}$ following a p-factor Bernoulli mixture model with default probabilities $Q_i(\boldsymbol{\psi})$ Assumptions A1 and A3 are automatically satisfied; A2 becomes

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

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 4.6. *Consider a sequence $L^{(m)} = \sum_{i=1}^{m} L_i$ satisfying Assumptions A1, A2 and A3 with a one-dimensional mixing variable Ψ with distribution function $G(\psi)$. 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*

$$\lim_{m\to\infty} \frac{1}{m}\, q_\alpha(L^{(m)}) = \overline{\ell}(q_\alpha(\Psi))\,. \tag{16}$$

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.

It follows from Proposition 4.6, that the tail of the credit loss in large one-factor Bernoulli mixture models is essentially driven by the tail of the mixing variable Ψ. Consider in particular 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 4.6 implies that for large m the tail of $M^{(m)}$ is heavier in model 1 than in model 2.

Proposition 4.6 shows that in a large credit portfolio with losses following a one factor Bernoulli mixture model the quantile

$$q_\alpha(L^{(m)}) = m\, q_\alpha(m^{-1}L^{(m)}) \approx m\, \overline{\ell}(q_\alpha(\Psi))$$

grows linearly in the size of the portfolio, so that there are no further diversification effects taking place when we increase the portfolio. This can be taken as a justification for the capital adequacy rule in the internal ratings based approach of Basel II; see Gordy (2001) for an interesting discussion of this point.

Example 4.7. 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 (16) implies for large m

$$q_\alpha(L^{(m)}) \approx m \sum_{r=1}^{k} \lambda_r h(\mu_r + \sigma_r \Phi^{-1}(\alpha)) \,. \tag{17}$$

Example 4.8. In this example we use Proposition 4.6 to study the model risk related to different specifications of the mixing-distribution in various exchangeable Bernoulli mixture models, assuming that default probability π and default-correlation ρ_Y (or equivalently π and π_2) are known and fixed for all models. As explained above the tail of M is essentially determined by the tail of the mixing distribution $G(q)$. In Figure 1 we plot the tail function of the probit-normal mixing model, the logit-normal model, the beta-model and the mixing model corresponding to the Clayton copula (see Example 4.14 below) on a logarithmic scale. 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 the estimation of these parameters is a difficult task, and we will discuss several approaches in Section 5 below.

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) for the special case of CreditMetrics and CreditRisk$^+$, 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 4.9. 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 random variables $(X_i)_{1\le i\le m}$ are independent.

Proposition 4.10. *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 $\mathbf{\Psi}$.

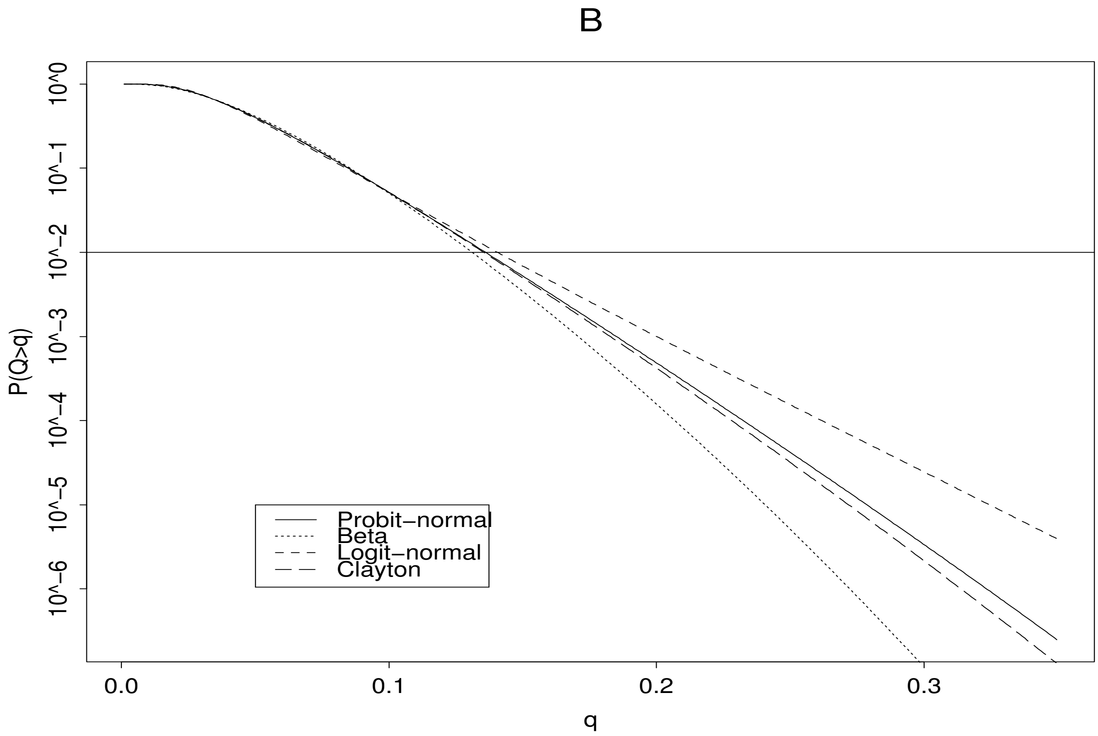

Figure 1: Tail of the mixing distribution G of Q in four different exchangeable Bernoulli mixture models: beta (close to one-factor CreditRisk$^+$); probit-normal (one-factor KMV/CreditMetrics); logit-normal (CreditPortfolioView); Clayton. In all cases the first two moments π and π_2 have the values for group B in Table 1. Horizontal line at 0.01 shows that models only really start to diverge around 99th percentile of mixing distribution.

(ii) *For any choice of thresholds d_1^i, $1 \le i \le m$ the default indicators $Y_i = 1_{\{X_i \le d_1^i\}}$ follow a Bernoulli mixture model with factor $\mathbf{\Psi}$; the conditional default probabilities are given by $Q_i(\mathbf{\Psi}) = P(X_i \le d_1^i \mid \mathbf{\Psi})$.*

Example 4.11 (Normal mean-variance mixtures with factor structure). Suppose that the latent variables $\mathbf{X}$ have a normal mean-variance mixture distribution as in Example 3.5 so that $X_i = \boldsymbol{\mu}_i(W) + g(W)Z_i$ for W independent of Z_i. Suppose also that $\mathbf{Z}$ (and hence $\mathbf{X}$) follows the linear factor model (3), so that $Z_i = \sum_{j=1}^{p} a_{i,j}\Theta_j + \sigma_i\varepsilon_i$ for a random vector $\mathbf{\Theta} \sim N_p(\mathbf{0}, \Omega)$ and independent, standard normally distributed random variables $\varepsilon_1 \ldots, \varepsilon_m$, which are also independent of $\mathbf{\Theta}$. Then $\mathbf{X}$ has a $(p+1)$-dimensional conditional independence structure.

To see this define the $(p+1)$-dimensional random vector $\mathbf{\Psi}$ by $\mathbf{\Psi} = (\Theta_1, \ldots, \Theta_p, W)'$ and observe that conditional on $\mathbf{\Psi}$ the random variables X_i are 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 $(d_1^i)_{1\le i\le m}$ we get conditional default probabilities

$$Q_i(\mathbf{\Psi}) = P(X_i \le d_1^i \mid \mathbf{\Psi}) = \Phi\left(\frac{d_1^i - \mu_i(W) - g(W)\sum_{j=1}^{p} a_{ij}\Theta_j}{g(W)\sigma_i}\right). \tag{18}$$

In the special case of multivariate t latent variables we obtain

$$Q_i(\mathbf{\Psi}) = \Phi\left(\sigma_i^{-1}\left(d_1^i\sqrt{W/\nu} - \sum\nolimits_{j=1}^{p} a_{ij}\Theta_j\right)\right). \tag{19}$$

The formula (18) is the key to Monte Carlo simulation for latent variable models with normal mixture distributions in a large portfolio context. For example, rather than simulating an m-dimensional t distribution to implement the Student model, we are simply required to simulate a p-dimensional normal vector $\mathbf{\Theta}$ with $p \ll m$, an independent chi-squared variate W and then to conduct a series of independent Bernoulli experiments with default probabilities $Q_i(\mathbf{\Psi})$ to decide whether individual counterparties default.

Example 4.12 (Archimedean copulas). As shown in the following lemma, which is essentially due to Marshall and Olkin (1988), latent variable models based on exchangeable Archimedean copulas possess a one-dimensional conditional independence structure. The relevance of this result to credit risk modelling is also discussed in Schönbucher (2002).

Lemma 4.13. *Given a distribution function 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 random variable $\Psi \sim F$ and a sequence $(U_i)_{1\le i\le m}$ of random variables which are conditionally independent given Ψ with conditional distribution function $P(U_i \le u \mid \Psi = \psi) = \exp(-\psi\varphi^{-1}(u))$ for $u \in [0,1]$. Then*

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

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

The Lemma gives a recipe for simulating from an Archimedean copula with generator ϕ. We need to find a distribution function whose Laplace transform is ϕ^{-1} so that we can simulate values of Ψ. In a second stage we then simulate independently variates U_i with distribution function $F(u) = \exp(-\Psi\phi(u))$. For a list of some Archimedean copulas where this is possible see Joe (1997) or Schönbucher (2002).

Consider now a latent variable model (X_i, d_1^i), $1 \le i \le m$ where $\mathbf{X}$ has an exchangeable Archimedean copula with generator ϕ. Put $Y_i = 1_{\{X_i \le d_1^i\}}$ and $\overline{p}_i = P(Y_i = 1)$. Using Lemma 4.13, an equivalent Bernoulli mixture model is now straightforward to compute.

Observe that for Ψ and $U_1 \dots U_m$ as in the Lemma $\left(X_i, d_1^i\right)_{1 \le i \le m}$ 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 are obviously independent given Ψ and we obtain for the conditional default probabilities

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

To simulate from the Archimedean copula-based latent variable model we may therefore use the following efficient and simple approach. In a first step we simulate a realisation of Ψ and then we conduct m independent Bernoulli experiments with default probabilities $Q_i(\Psi)$ to simulate a realisation of defaulting counterparties.

Example 4.14 (The Clayton Copula). As a concrete example again consider the Clayton copula of Example 3.6 with generator $\phi(t) = t^{-\theta} - 1$. Suppose we wish to construct an exchangeable Bernoulli mixture model with default probability π and joint default probability π_2 which is equivalent to a latent variable model driven by the Clayton copula. Using (4) the required value of θ to give the desired default probabilities is the solution to the equation

$$\pi_2 = C_\theta(\pi, \pi) = (2\pi^{-\theta} - 1)^{-1/\theta}, \ \ \theta > 0.$$

It is easily seen that π_2 and hence default correlation in our exchangeable Bernoulli mixture model is increasing in θ; for $\theta \to 0$ we obtain independent defaults, for $\theta \to \infty$ defaults become comonotonic and default correlation tends to one. A gamma$(1/\theta)$ variate Ψ with density $g(q) = q^{1/\theta - 1} \exp(-q)/\Gamma(1/\theta)$ has Laplace transform equal to the generator inverse $\phi^{-1}(t) = (t + 1)^{-1/\theta}$, so that a mixing distribution on $[0, 1]$ would be defined by setting $Q = \exp(-\Psi(\pi^{-\theta} - 1))$. In effect we use a mixing distribution where $-\log Q$ has a two-parameter gamma distribution. The Bernoulli mixture model implied by the Clayton copula will be used in Section 5.1. See also Schönbucher (2002) for more discussion of the technique used in this example.

5 Calibration of Bernoulli mixture models

In this section we consider fitting Bernoulli mixture models to historical default data. We envisage three situations of increasing complexity.

- Calibration of a model for a single homogeneous group of obligors with some common credit rating. We consider various approaches to estimating the parameters of an exchangeable Bernoulli mixture model, in particular the default correlation.

- Calibration of a model for a large portfolio divided into a number of different rating categories. Here we assume that rating category is the only available covariate for each obligor and that historical default data have been collected for each rating class (either internally or by a rating agency using a comparable rating system). We fit the model of Example 4.4 to data collected in Standard and Poor's (2001).

- Calibration of a latent variable model where the latent variables have a normal variance mixture distribution. Here we envisage a portfolio where the default potential of individual obligors is considered to be better understood so that they are treated more heterogeneously. We assume that a partial calibration of the model is undertaken using the approach of KMV/CreditMetrics but that some parameters of the latent variable distribution remain unknown and these are to be estimated from historical data using the equivalent Bernoulli mixture model.

5.1 Calibration of an exchangeable model

Suppose we have n years of data on historical default numbers for a homogeneous group; for $j = 1, \ldots, n$ let m_j denote the number of obligors observed in year j and let M_j denote the number that default. Further suppose that these defaults are generated by an exchangeable Bernoulli mixture model so that there exist identically distributed mixing variables $Q_1, \ldots, Q_n$ and defaults in year j are conditionally independent given Q_j. We consider two generic methods for estimating the fundamental parameters $\pi = \pi_1$, π_2 and ρ_Y: the method of moments and the maximum likelihood method.

A natural moment-style estimator of π_k is given by

$$\widehat{\pi}_k = \frac{1}{n}\sum_{j=1}^{n} \frac{\binom{M_j}{k}}{\binom{m_j}{k}} = \frac{1}{n}\sum_{j=1}^{n} \frac{M_j(M_j-1)\cdots(M_j-k+1)}{m_j(m_j-1)\cdots(m_j-k+1)}, \quad 1 \leq k \leq \min\{m_1, \ldots, m_n\}. \tag{20}$$

To understand this estimator observe that $\binom{M_j}{k}$ represents the number of possible subgroups of k obligors among the defaulting obligors in year j. If we write $Y_{j,1}, \ldots, Y_{j,m_j}$ for the default indicators of the obligors oberved in year j we have

$$\binom{M_j}{k} = \sum_{i_1,\ldots,i_k \colon \{i_1,\ldots,i_k\} \subset \{1,\ldots,m_j\}} Y_{j,i_1} \cdots Y_{j,i_k},$$

so that $E(\binom{M_j}{k}/\binom{m_j}{k}) = \pi_k$ follows by taking expectations of both sides. We estimate the unknown theoretical moment π_k by taking the natural empirical average (20) constructed from the n years of data. The estimator is unbiased for π_k and consistent (as $n \to \infty$); for more details see Frey and McNeil (2001). Obviously ρ_Y can be estimated by taking $\widehat{\rho}_Y = (\widehat{\pi}_2 - \widehat{\pi}^2)/(\widehat{\pi} - \widehat{\pi}^2)$.

An alternative moment-style estimator of π_2 has been proposed by Gordy (2000) and, in the notation of our paper, this takes the form

$$\widetilde{\pi}_2 = \widehat{\pi}^2 + \frac{\frac{1}{n}\sum_{j=1}^{n}\left(\frac{M_j}{m_j} - \widehat{\pi}\right)^2 - \frac{\widehat{\pi}(1-\widehat{\pi})}{n}\sum_{j=1}^{n}\frac{1}{m_j}}{1 - \frac{1}{n}\sum_{j=1}^{n}\frac{1}{m_j}}. \tag{21}$$

For a derivation see Appendix B of Gordy (2000).

To implement a maximum likelihood (ML) procedure we assume a simple parametric form for the density of the Q_j (such as beta, logit- or probit-normal); the joint probability function of the data is then calculated using (10) under the assumption that the Q_j are independent and maximised with respect to the natural parameters of the mixing distribution (i.e. a and b in the case of beta and μ and σ for the logit- and probit-normal). If independence seems unrealistic the method can be considered as a quasi-maximum-likelihood (QML) procedure which should still yield reasonable parameter estimates. The ML estimates of $\pi = \pi_1$, π_2 and ρ_Y are calculated by evaluating moments of the fitted distribution using (11).

To decide which of these approaches is the best way to estimate these parameters we have conducted a simulation study (summarised in Table 2) of the performance of the moment method using both (20) and (21) and the ML method based on the assumption of a beta mixing distribution. Computationally the beta mixture model is the easiest of the mixture models to fit because evaluation of the joint probability of the data can be achieved without numerical integration; this makes it fast and easy to perform in a large simulation study.

To generate data in the simulation study we consider three different Bernoulli mixture models: the beta and probit-normal mixing models of Section 4.1.1; the mixing model implied by a latent variable model with Clayton copula as in Example 4.14. In any single experiment we generate 20 years of data using parameter values that roughly correspond

Model	Parameter	CCC	B	BB
All models	π	0.188	0.049	0.0112
	π_2	0.042	0.00313	0.000197
	ρ_Y	0.0446	0.0157	0.00643
Beta	a	4.02	3.08	1.73
	b	17.4	59.8	153
Probit-Normal	μ	-0.93	-1.71	-2.37
	σ	0.316	0.264	0.272
Clayton	π	0.188	0.049	0.0112
	θ	0.0704	0.032	0.0247

Table 1: Parameter values used in the simulation study of Table 2. These correspond very roughly to the CCC, B and BB rating classes of Standard and Poors.

to one of the credit ratings CCC, B or BB of Standard & Poors; see Table 1 for the parameter values. The number of firms m_j in each of the years is generated randomly using a binomial-beta model to give a spread of values typical of real data; the defaults are then generated using one of the Bernoulli mixture models and the three methods are compared. The experiment is repeated 5000 times and a relative root mean square error (RRMSE) is estimated for each parameter and each method; that is we take the square root of the estimated MSE and divide by the true parameter value.

It may be concluded from Table 2 that for estimating default probabilities the standard method (20) is good enough and gives an essentially identical performance to an approach based on fitting a Bernoulli mixture model with beta mixing distribution. However for estimating π_2 and ρ_Y the ML method is (almost) alway best and outperforms the moment estimators even in the models where the beta distribution is misspecified. This can be explained by the fact that - as shown in Example 4.8 - when we constrain well-behaved, unimodal mixing distributions with densities (such as our four choices) to have the same first and second moments these distributions are very similar. Of the moment formulae for π_2 (20) should be preferred as the better quick method of getting an estimate. It can also be noted that the ML method tends to outperform the moment methods more as we increase the credit quality so that defaults become rarer.

5.2 Calibration of model for several exchangeable groups

In view of the results in Table 2 we apply a ML approach to fitting more complex models. We now suppose that in each year we have data 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.

It is possible to generalise the beta mixture model of the previous section to obtain a regression model for grouped data (see Joe, page 216), but numerical integration to obtain the model likelihood can generally not be avoided in realistic models and we find it equally convenient to switch to the probit-normal (or logit-normal) mixing distribution and fit the model of Example 4.4. Thus 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, and the unconditional probability function is obtained by integrating over this factor.

To complete the specification we assume that the factor random variables $\Psi_1, \ldots, \Psi_n$ for each of the different years are iid standard normal. Again this may be considered to be a QML approach if the assumption of independent factor variables seems unrealistic. 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 the marginal distributions of

Group	True Model	Par.	MomentA		MomentB		MLEbeta	
			RRMSE	Δ	RRMSE	Δ	RRMSE	Δ
CCC	beta	π	0.101	0	0.101	0	0.101	0
CCC	beta	π_2	0.202	0	0.204	1	0.201	0
CCC	beta	ρ_Y	0.332	5	0.344	9	0.317	0
CCC	clayton	π	0.102	0	0.102	0	0.102	0
CCC	clayton	π_2	0.205	1	0.207	2	0.204	0
CCC	clayton	ρ_Y	0.331	8	0.344	12	0.306	0
CCC	probitnorm	π	0.100	0	0.100	0	0.100	0
CCC	probitnorm	π_2	0.205	1	0.208	2	0.204	0
CCC	probitnorm	ρ_Y	0.347	11	0.361	15	0.314	0
B	beta	π	0.130	0	0.130	0	0.130	0
B	beta	π_2	0.270	0	0.275	2	0.269	0
B	beta	ρ_Y	0.396	8	0.409	12	0.367	0
B	clayton	π	0.134	0	0.134	0	0.133	0
B	clayton	π_2	0.293	3	0.299	5	0.284	0
B	clayton	ρ_Y	0.432	17	0.449	22	0.368	0
B	probitnorm	π	0.130	0	0.130	0	0.130	0
B	probitnorm	π_2	0.286	3	0.292	5	0.277	0
B	probitnorm	ρ_Y	0.434	19	0.451	24	0.364	0
BB	beta	π	0.199	0	0.199	0	0.199	0
BB	beta	π_2	0.435	0	0.447	3	0.438	1
BB	beta	ρ_Y	0.508	7	0.526	10	0.476	0
BB	clayton	π	0.196	0	0.196	0	0.195	0
BB	clayton	π_2	0.475	8	0.490	12	0.438	0
BB	clayton	ρ_Y	0.588	22	0.610	27	0.480	0
BB	probitnorm	π	0.197	0	0.197	0	0.197	0
BB	probitnorm	π_2	0.492	10	0.509	14	0.446	0
BB	probitnorm	ρ_Y	0.607	27	0.631	31	0.480	0

Table 2: Each panel of the table relates to a block of 5000 simulations using a particular exchangeable Bernoulli mixture model with parameter values roughly corresponding to a particular S&P rating class. For each parameter of interest an estimated RRMSE (relative root MSE) is tabulated for each of the three estimation methods: moment estimation based on (20) denoted MomentA; moment estimation based on (21) denoted MomentB; ML estimation based on the beta model. Methods can be compared by using Δ - the percentage inflation of the estimated RRMSE with respect to the best method (i.e. the RRMSE minimising method) for each parameter. Thus for each parameter the best method has $\Delta = 0$. The table clearly shows that MLE performs (almost) always best.

the $\mathbf{M}_j$ and the log-likelihood takes the form

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

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ and $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_k)'$ are the unknown model parameters and

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

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

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

Having obtained estimates of $\widehat{\boldsymbol{\mu}}$ 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. We use (13) to calculate estimates

$$\begin{aligned} \widehat{\pi}^{(r)} &:= \int_{-\infty}^{\infty} h(\widehat{\mu}_r + \widehat{\sigma}_r z)\phi(z)dz, \qquad 1 \le r \le k \\ \widehat{\pi}_2^{(r,s)} &:= \int_0^1 \left(h(\widehat{\mu}_r + \widehat{\sigma}_r \Phi^{-1}(q))h(\widehat{\mu}_s + \widehat{\sigma}_s \Phi^{-1}(q))\right) dq, \quad 1 \le r, s \le k, \end{aligned}$$

where $\widehat{\pi}_2^{(r,s)}$ gives the estimated default probability for a pairs of obligors chosen respectively from groups r and s. The matrix of estimated within-group and between-group default correlations has (r, s)-element given by

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

where the diagonal elements $\widehat{\rho}_Y^{(r,r)}$ are the estimated within-group default correlations.

Example 5.1 (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 3, together with the estimated default probabilities $\widehat{\pi}^{(r)}$ and estimated default correlations $\widehat{\rho}_Y^{(r,s)}$ that these imply. These values may prove a useful reference for researchers who seek plausible values for default correlations and probabilities in other studies of portfolio credit risk. Note that these default correlations are correlations between event indicators for very low probability events and are necessarily very small.

The fitted model summarised in Table 3 can be used in conjunction with (17) to give large sample risk measure estimates. The steps required are as follows.

1. The portfolio is mapped to the S&P rating system. For example, consider a low-grade portfolio of 10000 obligors where the numbers of A, BBB, BB, B and CCC-rated firms are 2000, 1000, 1000, 3000, 3000.

Parameter	A	BBB	BB	B	C
μ_r	-3.40	-2.90	-2.41	-1.69	-0.84
s.e.(μ_r)	0.14	0.09	0.08	0.06	0.08
σ_r	0.189	0.205	0.252	0.239	0.262
s.e.(σ_r)	0.17	0.10	0.07	0.05	0.07
$\pi^{(r)}$	0.004	0.0022	0.0098	0.0503	0.2066
$\rho_Y^{(r,s)}$	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: 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 implied estimates of default probabilities $\widehat{\pi}^{(r)}$ and default correlations $\widehat{\rho}_Y^{(r,s)}$ that these imply. Note that we have tabulated default correlation in absolute terms and not in percentage terms.

2. The inputs to formula (17) are determined. In our case we have $m = 10000$, $k = 5$, $\lambda_1 = 0.2, \lambda_2 = 0.1, \lambda_3 = 0.1, \lambda_4 = 0.3, \lambda_5 = 0.3$. The parameters $\mu_1, \ldots, \mu_5, \sigma_1, \ldots, \sigma_5$ are replaced by the estimates in Table 3. The h function is the probit link function $h(x) = \Phi(x)$, wher Φ is the standard normal distribution function.

3. For typical values of α, such as 99% or 99.9% the formula is evaluated to give estimates of the corresponding credit portfolio VaR. In our example we get $q_{0.99}(L^{(10000)}) \approx 1652$ and $q_{0.999}(L^{(10000)}) \approx 2039$.

The implied default probability and default correlation estimates in Table 3 could be used to calibrate stochastic models other than the probit-normal model to the S&P default data. For example, to calibrate a Clayton copula to group BB we use the inputs $\pi^{(3)} = 0.0098$ and $\rho_Y^{(3,3)} = 0.00484$ to determine the parameter θ of the Clayton copula.

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 estimate all parameters of an equivalent mixture model from historical default data practically impossible, since relevant data for many parameters 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.[1] For a Gaussian latent variable model this completes the calibration but, as we have seen in Example 3.5, it is possible to develop latent variable models with non-Gaussian

[1] Note that this is exactly the information which is provided by the KMV-Moodys model to subscribers of the service.

distribution and more parameters. In this section we will consider how the calibration of a model with multivariate t-distributed latent variables could be completed by estimating the degrees of freedom parameter ν; thus we obtain a hybrid calibration method 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. The approach would extend to other normal mean-variance mixture models with more unknown parameters in the mixing distribution.

As before let m_j, $1 \leq j \leq n$ be the number of obligors in the sample of observed firms in year j. In any given year the default indicator vector $\mathbf{Y_j} = (Y_{j,1}, \ldots, Y_{j,m_j})'$ is induced by a latent variable model $(X_{j,i}, d_1^{j,i})$, $1 \leq i \leq m_j$, where the latent variable vector $\mathbf{X}_j$ follows a multivariate t distribution with factor structure. The conditional default probability of the ith obligor in year j follows from (19) and is given by

$$P(Y_{j,i} = 1 \mid \boldsymbol{\Theta_j} = \boldsymbol{\theta}, W_j = w) = Q_{j,i}(\boldsymbol{\theta}, w) := \Phi\left(\frac{d_1^{j,i}\sqrt{w/\nu} - (A_j\boldsymbol{\theta})_i}{\sigma_{j,i}}\right), \qquad (22)$$

where in every year j the matrix A_j (determining the factor structure) and the constants $d_1^{j,i}$ and $\sigma_{j,i}$ (determining the individual default probabilities) are considered known. Furthermore we assume that W_j is independent of $\boldsymbol{\Theta_j}$ for all j and that $(W_j, \boldsymbol{\Theta_j})_{1\leq j\leq n}$ forms an iid sequence of random vectors.

To complete the model fitting we use ML estimation to determine the parameter ν of the chi-squared distribution of the W_j. Denote by $\mathbf{y}_j = (y_{j,1}, \ldots, y_{j,m_j})'$ the default-observations in year j. Let $B_j = \{1 \leq i \leq m_j : y_{j,i} = 1\}$ be the identities of the firms which have defaulted in year j, and $B_j^c := \{1, \ldots, m_j\} - B_j$ the identities of the surviving firms. Using the conditional independence of the default indicators we obtain

$$P(\mathbf{Y}_j = \mathbf{y}_j \mid W_j, \boldsymbol{\Theta}_j) = \prod_{i\in B_j} P(Y_{j,i} = 1 \mid \boldsymbol{\Theta_j}, W_j) \prod_{i\in B_j^c} (1 - P(Y_{j,i} = 1 \mid \boldsymbol{\Theta_j}, W_j)). \qquad (23)$$

The unconditional default probability function of the observations depends on the unknown parameter ν and is given by

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

where $F_{\mathbf{0},\Omega_j}(\boldsymbol{\theta})$ denotes the joint distribution function of a $N_p(\mathbf{0}, \Omega_j)$-distributed random vector and G_ν denotes the distribution function of a χ^2_ν random variable. Under our independence assumptions the log-likelihood function for given observations $\mathbf{y}_1, \ldots, \mathbf{y}_n$ equals $L(\boldsymbol{\nu}; \mathbf{y}_1, \ldots, \mathbf{y}_n) = \sum_{j=1}^n \ln P(\mathbf{Y}_j = \mathbf{y}_j\,; \nu)$, and this is maximised with respect to ν.

The main practical obstacle is the evaluation of the double integral in the log-likelihood and the approach that we use is Monte Carlo simulation. We make the following practical observations about the method. First, although the likelihood contribution (23) assumes that every obligor is completely heterogeneous it is advisable to group obligors as far as possible into homogeneous groups with identical default probabilities and factor structure. The main advantage of this approach is better numerical performance of the MC-simulation in the computation of the log-likelihood function. Second, it is very useful to use multivariate importance sampling. 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

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

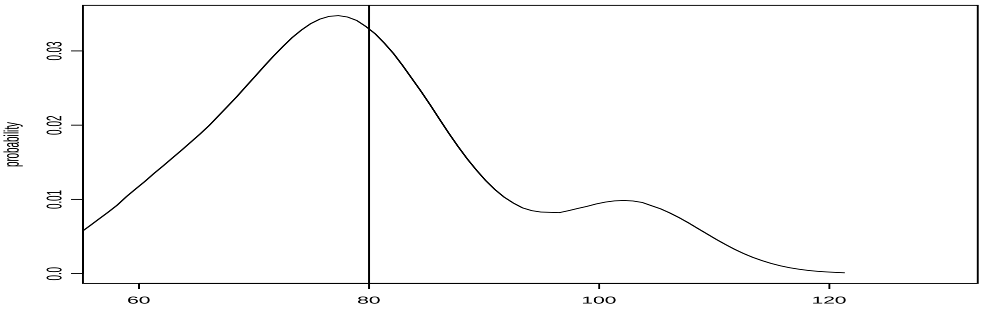

Figure 2: Density plot of estimates of the 99% quantile of M, the number of defaults, in a latent variable model with t distributed latent variables and factor structure. The true value is approximately equal to 80. Details are given in Section 5.3.

The new mean $\boldsymbol{\mu}(w)$ of the relocated multivariate normal distribution with distribution function $F_{\boldsymbol{\mu}(w),I}$, 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 random variables; in particular the density ratio $dF_{\mathbf{0},I}/dF_{\boldsymbol{\mu}(w),I}$ is easily computed.

Example 5.2 (A simulation Experiment). To assess the feasibility of the above approach in real applications we constructed a hypothetical situation using simulated data. We assumed the availability of 20 years of data and considered a test portfolio of 400 obligors belonging to three homogeneous groups with different exposures to two systematic factors. The latent variables governing default were assumed to have a t distribution with 10 degrees of freedom. $K = 100$ random datasets of 20-year default patterns were generated and ν was estimated from each dataset using the MC approach to evaluating the likelihood.

For each estimate $\widehat{\nu}_k, k = 1, \ldots, K$ we evaluated $q_{0.99}(M; \widehat{\nu}_k)$, the approximate 99% quantile of the distribution of defaults implied by (16). Since the tail of the loss distribution is the key object of interest in credit portfolio management, a density plot of the estimated values $q_{0.99}(M; \hat{\nu}_k)$, $k = 1, \ldots, K$ as in Figure 2 permits a good assessment of the performance of the estimator. The true value in our simulated model is $q_{0.99}(M; 10) \approx 80$ and in a Gaussian model it would be $q_{0.99}(M; \infty) \approx 58$. Since our estimates are concentrated around 80 it appears that the estimator is mostly able to distinguish the riskier t model with $\nu = 10$ from a normal model.[2] This suggests that in the context of a normal variance mixture models ML estimation might be useful in reducing the model risk related to the choice of the copula.

[2]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 used in the density plot.

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 summarise the practical aspects of our research.

Bernoulli mixture models are more convenient than latent variable models when it comes to both estimation and simulation, at least for large portfolios. If 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 a 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 4.4 are reasonable models for portfolios with a relatively homogeneous exposure to a common set of risk factors. They are also useful parsimonious models in a situation where we have only rather imprecise information on the risk factors affecting a portfolio, such that we have to rely solely on historical default information in estimating model parameters. As we have seen in Section 5.1 and Section 5.2, maximum likelihood fitting of Bernoulli mixture models is a feasible method of obtaining estimates of model parameters and the precise choice of the mixing distribution is of lesser importance. Moreover, maximum likelihood methods seem to perform better than simple moment estimation techniques. We further note that the models we have fitted essentially belong to the class of generalised linear mixed models (GLMMs); although we have taken a maximum likelihood approach it would also be possible to use computational Bayesian methods such as Markov Chain Monte Carlo to fit such models.

For a portfolio, where obligors are exposed to several risk factors and where the exposure to different risk factors differs markedly across 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.

References

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

Crosbie, P., and J. Bohn (2002): "Modeling default risk," KMV working paper, available from `http://www.kmv.com`.

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," ETH E-Collection, URL http://e-collection.ethbib.ethz.ch/show?type=bericht&nr=85, ETH Zürich.

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*, pp. 1317–1344.

FREY, R., A. MCNEIL, and N. NYFELER (2001): "Copulas and Credit Models," *RISK*,14(October) pp. 111–114.

GIESECKE, K. (2001): "Structural modelling of defaults with incomplete information," preprint, Humboldt-Universität Berlin, forthcoming in Journal of Banking and Finance.

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, forthcoming in Journal of Financial Intermediation.

JARROW, R., and F. YU (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.

KEALHOFER, S., and J. BOHN (2001): "Portfolio management of default risk," KMV working paper, available from `http://www.kmv.com`.

KOYLUOGLU, U., and A. HICKMAN (1998): "Reconciling the Differences," *RISK*, 11(10), 56–62.

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 I. OLKIN (1988): "Families of multivariate distributions," *Journal of the American Statistical Assosiation*, 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., and D. SCHUBERT (2001): "Copula-dependent default risk in intensity models," preprint, Universität Bonn.

SCHÖNBUCHER, P. J. (2002): "Taken to the limit," Preprint, 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 distribution function of such a distribution.

We use the notation $C(\mathbf{u}) = C(u_1, \ldots, u_d)$ for the d-dimensional joint distribution functions 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. $C(u_1, \ldots, u_d)$ is increasing in each component u_i.
2. $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ for all $i \in \{1, \ldots, d\}$, $u_i \in [0,1]$.
3. For all $(a_1, \ldots, a_d), (b_1, \ldots, b_d) \in [0,1]^d$ with $a_i \le b_i$ we have:

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

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

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

$$\begin{aligned} F(x_1, \ldots, x_n) &= P\left(F_1(X_1) \le F_1(x_1), \ldots, F_d(X_d) \le F_d(x_d)\right) \\ &= C(F_1(x_1), \ldots, F_d(x_d)), \end{aligned} \tag{24}$$

or alternatively $C(u_1, \ldots, u_n) = F\left(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d)\right)$, where $F_i^{\leftarrow}$ denotes the generalised inverse of the distribution function F_i. Formula (24) 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 $F_1, \ldots, F_d$. Then there exists a copula $C : [0,1]^d \to [0,1]$ such that for all $x_1, \ldots, x_d$ in $\overline{\mathbb{R}} = [-\infty, \infty]$ (24) holds; C is unique if $F_1, \ldots, F_d$ are continuous. Conversely, if C is a copula and $F_1, \ldots, F_d$ are distribution functions, then the function F given by (24) is a joint distribution function with margins $F_1, \ldots, F_d$.*

For a proof we refer to Schweizer and Sklar (1983). If F is a joint distribution function with marginals $F_1, \ldots, F_d$ and (24) 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 $(X_1, \ldots, X_d)$ be a vector of continuously distributed risks with copula C and let $T_1, \ldots, T_d$ be strictly increasing functions. Then it is easily seen that $(T_1(X_1), \ldots, T_d(X_d))$ 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

$$C_R^{\mathrm{Ga}}(\mathbf{u}) = \mathbf{\Phi}_R\left(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)\right), \tag{25}$$

where $\mathbf{\Phi}_R$ denotes the joint distribution function of a standard d-dimensional normal random vector $\mathbf{X}$ with correlation matrix R, and Φ is the distribution function of univariate standard normal.

B Proofs of Lemmas and Propositions

B.1 Lemma 3.2

Proof. For notational simplicity consider the case $m = 2$. Denote by C the copula of $\mathbf{X}$ and recall the following identity (see (24) 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 $u_{i,j} := P\left(X_i \le d_j^i\right) = P\left(\widetilde{X}_i \le \widetilde{d}_j^i\right)$, $j \in \{1, \dots, n\}$, $i = 1, 2$. Hence we get

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

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 $a_i \le b_i$, $i = 1, \dots, m$

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

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

B.2 Proposition 4.5

Proof. Our proof is based on a result of Petrov (see Petrov (1975), Theorem IX.12), which says that if $(Z_i)_{i\in\mathbb{N}}$ is a sequence of independent random variables with $E(Z_i) = 0$ and $\sum_{i=1}^{\infty} E(|Z_i|^2)/i^2 < \infty$, then $\frac{1}{m}\sum_{i=1}^{m} Z_i \to 0$ a.s.

To apply this result we put $Z_i := L_i - \ell_i(\boldsymbol{\psi})$. Since $0 \le \Delta_i Y_i \le 1$ we get

$$\sum_{i=1}^{\infty} E(Z_i^2)/i^2 = \sum_{i=1}^{\infty} e_i^2/i^2 \operatorname{var}(\Delta_i Y_i) \le \sum_{i=1}^{\infty} (e_i/i)^2,$$

which is finite by Assumption A3. □

B.3 Proposition 4.6

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

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

Since $\bar{\ell}$ 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

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

Since $\bar{\ell}$ 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 $m(\bar{\ell}(q_\alpha(\Psi)) - \varepsilon) \le q_\alpha(L^{(m)}) \le m(\bar{\ell}(q_\alpha(\Psi)) + \varepsilon)$. □

B.4 Proposition 4.10

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

$$P\left(\mathbf{Y} = \mathbf{y} \mid \mathbf{\Psi}\right) = P\left(\bigcap_{i\in B}\{X_i \le d_1^i\} \bigcap_{i\in B^c}\{X_i > d_1^i\} \mid \mathbf{\Psi}\right)$$
$$= \prod_{i\in B} P(X_i \le d_1^i \mid \mathbf{\Psi}) \prod_{i\in B^c}(1 - P(X_i \le d_1^i \mid \mathbf{\Psi})\,.$$

Hence conditional on $\mathbf{\Psi}$ the Y_i are independent Bernoulli variates with success-probability $Q_i(\mathbf{\Psi}) := P(X_i \le d_1^i \mid \mathbf{\Psi})$. The converse is obvious. □

B.5 Lemma 4.13

Proof. We have

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

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