Modelling Dependent Defaults

Rüdiger Frey*
Swiss Banking Institute
University of Zurich
Plattenstrasse 14
CH-8032 Zurich
Tel: +41 1 634 29 57
Fax: +41 1 634 49 03
freyr@isb.unizh.ch

Alexander J. McNeil*
Department of Mathematics
Federal Institute of Technology
ETH Zentrum
CH-8092 Zurich
Tel: +41 1 632 61 62
Fax: +41 1 632 15 23
mcneil@math.ethz.ch

August 13, 2001

Abstract

We consider the modelling of dependent defaults in large credit portfolios using latent variable models (the approach that underlies KMV and CreditMetrics) and mixture models (the approach underlying CreditRisk$^+$). We explore the role of copulas in the latent variable framework and show that for given default probabilities of individual obligors the distribution of the number of defaults in the portfolio is completely determined by the copula of the latent variables. We present results from a simulation study showing that, even for fixed asset correlations, assumptions concerning the latent variable copula can have a profound effect on the distribution of credit losses. In the mixture models defaults are conditionally independent given a set of common economic factors affecting all obligors and we explore the role of the mixing distribution of the factors in these models. In homogeneous, one-factor mixture models we find that the tail of the mixing distribution essentially determines the tail of the overall credit loss distribution. We discuss the relationship between latent variable models and mixture models and provide general conditions under which these models can be mapped into each other. Our contribution can be viewed as an analysis of the model risk associated with the modelling of dependence between individual default events.

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

Keywords: Portfolio Credit Risk Models, Model Risk, Dependence Modelling, Copulas, Mixture Models

1 Introduction

Introduction to credit-risk paper The development and analysis of quantitative models for credit losses in large lending portfolios has recently become a focus of attention for practitioners, regulators and academics. These models purport to capture the loss potential due to defaulting counterparties on the portfolio level; they are intended to be used for the measurement of the overall risk exposure in a large loan book, the active management of credit portfolios under risk-return considerations, or the pricing of credit insurance. Moreover, given improved availability of data on credit losses, refined versions of these 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.

*We wish to thank Dirk Tasche, Mark Nyfeler, Filip Lindskog, Uwe Schmock and Hans Bühlmann for helpful discussions.

A detailed description of the most popular credit risk models is given in Crouhy, Galai, and Mark (2000). The models currently in use can be divided into two classes. The models proposed by the KMV corporation (KMV-Corporation 1997) or the RiskMetrics group (RiskMetrics-Group 1997) are extensions of the firm-value model of Merton (1974). In these models, which are often termed latent variable models, default occurs if a latent variable, often interpreted as the value of the obligor's assets, falls below some threshold, often interpreted as the value of the obligor's liabilities. Dependence between defaults is caused by dependence between the latent variables. CreditRisk$^+$, developed by Credit Suisse Financial Products (Credit-Suisse-Financial-Products 1997), is, on the other hand, a typical actuarial model. In this model 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. The common factors causes defaults to be *unconditionally* dependent. In the statistics literature, such as Joe (1997), these models are referred to as mixture models.

Most of the previous research on credit risk concentrates on the modelling of individual defaults and the pricing of credit derivatives. In particular, researchers have developed sophisticated firm-value models extending the work of Merton (1974) and reduced form models for default of individual firms; see for instance Duffie (2000) for an overview. While these models can be extended to incorporate more than one counterparty – see for instance Duffie (1998) – portfolio considerations are not the main focus of this line of research. However, in the management of large, balanced, loan portfolios, the main risk for the lender is the occurrence of disproportionately many joint defaults of different counterparties over a fixed time horizon – in fact this is what might be termed "extreme credit risk" in this context.

A good model for the dependence of different obligors is essential to capture this type of extreme credit risk. In the present paper we consider the modelling of dependent defaults using recent conceptual insights on modelling dependence in risk management (see Embrechts, McNeil, and Straumann (2001)). In particular, the copula concept and the notion of extremal dependence of risk factors play a pivotal role in our analysis. Our first result clarifies the role of copulas in latent variable models. We show that for given default probabilities of individual firms the joint distribution of several defaults is completely determined by the copula of the latent variables. In this context it is easily seen that both the KMV and CreditMetrics are based on a Gaussian copula, as was observed independently by Li (1999). For multivariate-normally distributed risk factors (here the asset values of the different companies) the occurrence of many joint large movements of the risk factors is a rare event, since multivariate-normally distributed random variables are asymptotically independent. This casts some doubt on whether latent variable models based on a Gaussian copula are necessarily the best choice for modelling dependent defaults. To study this issue we replace the Gaussian copula by the copula of multivariate normal mean-variance mixtures such as the multivariate t or the multivariate hyperbolic distribution. We perform a simulation study which shows that this can have a drastic impact on the tail of the distribution of credit losses. It is well known that multivariate normal variance mixtures inherit the correlation matrix from the multivariate normal random vector used to construct them. Hence this shows that a proper fitting of the correlation matrix of the asset returns alone may not be sufficient to capture extreme credit risk. A short non-technical summary of this result can be found in Frey, McNeil, and Nyfeler (2001).

We go on to consider the modelling of dependent defaults in mixture models. We show that in this class of models the occurrence of extreme credit risk in large portfolios is closely related to the tail (or the moments) of the mixing distribution. This provides a theoretical underpinning for simulation results from Gordy (2000). We use simulations to compare the credit loss distribution for several popular mixing distributions, assuming that the first two moments of the mixing distribution, or equivalently default probability and default correlation, have been fixed. We also characterize analytically the worst-case mixing-distribution when the first two moments of the mixing distribution are given.

Moreover, we study the relationship between mixture models and latent variable models. We give conditions under which latent variable models and mixture models can be mapped into each other. Our results contain the results of Gordy (2000) on the relationship between the CreditMetrics and CreditRisk$^+$ models as a special case. Results on the connectios between the model types are of theoretical interest, as they help to clarify structural similarities and differences between the models. Moreover, they are of practical relevance, as they allow us to use simple simulation techniques for Bernoulli mixtures in the context of latent variable models.

Finally, we discuss a new estimator for joint default probabilities (moments of the mixing distribution) in exchangeable mixture models; these estimates are often very useful for calibrating models to historical default data from homogeneous obligor groups.

2 Models for Loan Portfolios

Consider a portfolio of m obligors and fix some time horizon T. For $1 \leq i \leq m$, let the random variable (rv) S_i be a state indicator for obligor i at time T. Assume that S_i takes integer values in the set $\{0, 1, \ldots, n\}$ representing for instance rating classes; we interpret the value 0 as default and non-zero values represent states of increasing creditworthiness. At time $t = 0$ obligors are assumed to be in some non-default state.

Often we will concentrate on the binary outcomes of default and non-default and ignore the finer categorization of non-defaulted companies. In this case we write Y_i for the default indicator variables; $Y_i = 1 \iff S_i = 0$ and $Y_i = 0 \iff S_i > 0$. The random vector $\mathbf{Y} = (Y_1, \ldots, Y_m)'$ is a vector of default indicators for the portfolio and

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

is its joint probability function; the marginal default probabilities will be denoted by $p_i = P(Y_i = 1)$, $i = 1, \ldots, m$.

We count the number of defaulted obligors at time $t = T$ with the random variable $M := \sum_{i=1}^m Y_i$. In the event of default the actual loss may also be modelled as a random quantity E_i, known as the loss given default. In practice this is assumed to be a random proportion of some known deterministic exposure. We will denote the overall loss by $L := \sum_{i=1}^m E_i Y_i$ and make further assumptions about the E_i's as and when we need them.

2.1 Exchangeable Models

To simplify the analysis we will often assume that the state indicator $\mathbf{S}$ and thus the default indicator $\mathbf{Y}$ is *exchangeable*. This seems the correct way to mathematically formalise the notion of *homogeneous* groups that is used in practice. Recall that a random vector $\mathbf{S}$ is called exchangeable if

$$(S_1, \ldots, S_m) \stackrel{\mathrm{d}}{=} (S_{\Pi(1)}, \ldots, S_{\Pi(m)}),$$

for any permutation $(\Pi(1), \ldots, \Pi(m))$ of $(1, \ldots, m)$. This implies in particular that for any $k \in \{1, \ldots, m-1\}$ 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.

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

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

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

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

It is important to distinguish π_k from $P(M = k)$. The distribution of M can however be calculated in terms of the π_k's in the following way.

$$\begin{aligned}
P(M = k) &= \binom{m}{k} P(Y_1 = 1, \ldots, Y_k = 1, Y_{k+1} = 0, \ldots, Y_m = 0) \\
&= \binom{m}{k} \sum_{S:\{1,\ldots,k\} \subset S \subset \{1,\ldots,m\}} (-1)^{|S|-k} \pi_{|S|} \\
&= \sum_{i=0}^{m-k} (-1)^i \frac{m!}{i!k!(m-k-i)!} \pi_{k+i}.
\end{aligned}$$

Existing approaches to modelling the distribution of the random vector $\mathbf{S}$ or $\mathbf{Y}$ belong essentially to two model classes - latent variable models and Bernoulli-mixture models. In the following sections we discuss each of these model types and explore the links between them.

3 Latent Variables Models

Definition 3.1. Let $\mathbf{X} = (X_1, \ldots, X_m)'$ be an m-dimensional random vector with continuous marginal distribution functions (df) $F_i(x) = P(X_i \leq x)$. For $i \in \{1, \ldots, m\}$ let

$$-\infty = D_{-1}^i < D_0^i < \cdots < D_n^i = \infty$$

be a sequence of *cut-off* levels. Set

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

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

X_i and D_0^i are often interpreted as the values of the assets i 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. Li (1999) works with another interpretation of the random vector $\mathbf{X}$ and takes X_i to be the time-to-default or survival time of company i; in his model company i defaults if $X_i < T$.

3.1 Industry examples: CreditMetrics and KMV

Structurally these models are quite similar; they differ with respect to the interpretation and calibration of the components of the model. In both models the latent vector $\mathbf{X}$ is assumed to have a multivariate normal distribution and X_i is interpreted as a change in asset value for obligor i over the time horizon of interest; D_0^i is chosen so that the probability of default for company i is the same as the historically observed default rate for companies of a similar credit quality. In CreditMetrics the classification of companies into groups of similar credit quality is generally based on an external rating system, such as that of Moodys or Standard & Poors. In KMV a *distance-to-default* is calculated for every company and companies with similar distances to default are grouped together. In both models the calibration of the correlation matrix is achieved by the use of a factor model relating changes to the latent variables to sytematic changes in a small number of underlying factors.

3.2 Latent variables and copulas

It is possible to set up different latent variable models which lead to the same multivariate distribution for $\mathbf{S}$; in particular, the distribution of $\mathbf{S}$ remains invariant under strictly increasing simultaneous transformation of the marginals X_i and the threshold values D_j^i. This prompts the idea of equivalence of latent variable models.

Definition 3.2. Let $\left(X_i, (D_j^i)_{-1\le j\le n}\right)_{1\le i\le m}$ and $\left(\widetilde{X}_i, (\widetilde{D}_j^i)_{-1\le j\le n}\right)_{1\le i\le m}$ be two latent variable models with state vectors $\mathbf{S}$ and $\widetilde{\mathbf{S}}$. The models are equivalent if $\mathbf{S} \stackrel{\mathrm{d}}{=} \widetilde{\mathbf{S}}$.

We now give a criterion for equivalence of two latent variable models in terms of the marginal distributions of the state vector $\mathbf{S}$ and the *copula* of $\mathbf{X}$. For more information on copulas we refer to Appendix A and to Embrechts, McNeil, and Straumann (2001).

Proposition 3.3. *Let* $\left(X_i, (D_j^i)_{-1\le j\le n}\right)_{1\le i\le m}$ *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*

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

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

2. $\mathbf{X}$ *and* $\widetilde{\mathbf{X}}$ *have the same copula.*

Proof. For notational simplicity consider the case $m = 2$. Denote by C the copula of $\mathbf{X}$ and recall the following identity (see (26) in Appendix A for more details).

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

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

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

For the case $m > 2$ we recall a useful identity. For all $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$ with $a_i \le b_i$, $i = 1,\dots,m$

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

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

□

Remark 3.4. The converse of the result is not generally true. If two latent variable models are equivalent then $\mathbf{X}$ and $\widetilde{\mathbf{X}}$ do not necessarily have the same copula.

This result underlines the structural equivalence of the CreditMetrics and KMV models. By assuming multivariate normality of the latent vectors both models work with the Gaussian copula. Moreover the model of Li (1999) can also be seen as equivalent to both CreditMetrics and KMV since the exponentially distributed survival times in this model are joined together by a Gaussian copula to form a multivariate distribution with exponential margins. There is no reason why we have to work with a Gaussian copula. In fact, the simulations in Section 6 show that the choice of copula may be very critical to the tail of the overall loss distribution, particularly when we consider relatively large portfolios of credit risks.

We now give a theoretical explanation why the copula of $\mathbf{X}$ is of relevance for the lower tail of the loss distribution. For simplicity we restrict ourselves to a model with only two states (default and non-default). Consider a subgroup of k companies $\{i_1, \dots, i_k\} \subset \{1, \dots, m\}$, with individual default probabilities $p_{i_1}, \dots, p_{i_k}$. Then

$$P(Y_{i_1} = 1, \dots, Y_{i_k} = 1) = P\left(X_{i_1} \le D_0^{i_1}, \dots, X_{i_k} \le D_0^{i_k}\right) = C_{i_1,\dots,i_k}(p_{i_1}, \dots, p_{i_k}), \qquad (2)$$

where $C_{i_1,\dots,i_k}$ denotes a k-dimensional margin of C. If $\mathbf{X}$ has an exchangeable copula (i.e. the copula of an exchangeable uniform random vector) and if all individual default probabilities are equal to some constant π, $\mathbf{Y}$ is exchangeable and (2) reduces to the useful formula

$$\pi_k = C_{1,\dots,k}(\underbrace{\pi, \dots, \pi}_{k \text{ times}}) \quad 1 \le k \le m. \qquad (3)$$

It is obvious from (2) and (3) that joint default probabilities for groups of obligors depend critically on the nature of the copula of the latent variables. Since individual default probabilities are typically small, copulas which have a more pronounced tendency to produce small values in several margins *simultaneously* than the Gaussian copula are of particular interest here, as they will lead to more joint defaults and hence to an increased likelihood of large aggregate losses. Such copulas are said to exhibit *lower tail dependence*; see Appendix A for a formal definition.

3.3 Latent variable models with non-Gaussian dependence structure

In light of the preceeding discussion it is natural to generalize the existing credit risk models by replacing the Gaussian copula with an alternative copula which exhibits lower tail dependence. There are various methods of constructing general m-dimensional copulas and useful references are Joe (1997), Nelsen (1999) and Lindskog (2000).

In this paper we concentrate on the copulas that are implicit when we assume multivariate normal (variance) mixture distributions for the latent variables. Our main example is the multivariate t-distribution. In contrast to the Gaussian copula, the t-copula has been shown to possess tail dependence in both tails; see Embrechts, McNeil, and Straumann (2001). Moreover, multivariate normal mixtures like the multivariate t or the generalized hyperbolic distribution are popular alternative models for financial returns and it is natural to investigate what happens when we use them in latent variable models for default and migration. A more formal discussion of this class of models is given in Section 5.2 below.

As an alternative we could use parametric copulas in closed-form. An example is provided by the class of so-called Archimedian copulas. The copulas in this class suffer from the deficiency that they are not rich in parameters and cannot model a fully flexible dependence between the latent variables. They can, however, model exchangeable or partially exchangeable random vectors, which may be sufficient for some latent variable models. The

Clayton copula provides an example with lower tail dependence. See Section 5.3 for an example.

4 Mixture Models

In a mixture model the default probability of an obligor is assumed to depend on a (typically small) 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 $\mathbf{\Psi} = (\Psi_1, \dots, \Psi_p)$, the random vector $\mathbf{Y} = (Y_1, \dots, Y_m)'$ 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 $P(Y_i = 1|\mathbf{\Psi}) = Q_i(\mathbf{\Psi})$.

Since default is a rare event we also explore the idea of approximating Bernoulli random variables with Poisson random variables in Poisson mixture models. Here a company is allowed to default more than once, albeit this is typically a low-probability event; the state indicator $\widetilde{Y}_i \in \{0, 1, 2, \dots\}$ gives the number of defaults of company i. The formal definition parallels the definition of a Bernoulli-mixture model.

Definition 4.2 (Poisson Mixture Model). Given p and $\mathbf{\Psi}$ as in Definition 4.1, the random vector $\widetilde{\mathbf{Y}} = (\widetilde{Y}_1, \dots, \widetilde{Y}_m)'$ follows a Poisson mixture model with factors $\mathbf{\Psi}$, if there are functions $\Lambda_i : \mathbb{R}^p \to (0, \infty)$, $1 \leq i \leq m$, such that conditional on $\mathbf{\Psi}$ the random vector $\widetilde{\mathbf{Y}}$ is a vector of independent Poisson$(\Lambda_i(\mathbf{\Psi}))$-distributed random variables.

Suppose that $\widetilde{\mathbf{Y}}$ has a Poisson mixture model. If we define the indicators $Y_i = 1_{\{\widetilde{Y}_i \geq 1\}}$ then $\mathbf{Y}$ follows a Bernoulli mixture model and the mixing variables are related by $Q_i(\cdot) = 1 - \exp(-\Lambda_i(\cdot))$. For small default-intensities Λ_i the random variable $\widetilde{M} = \sum_{i=1}^m \widetilde{Y}_i$ is approximately equal to the number of defaulting companies. Note that $\widetilde{M}$ is conditionally Poisson given $\mathbf{\Psi}$ with parameter $\bar{\Lambda} := \sum_{i=1}^m \Lambda_i(\mathbf{\Psi})$ so that

$$P(\widetilde{M} = k \mid \mathbf{\Psi}) = e^{-\bar{\Lambda}} \frac{\bar{\Lambda}^k}{k!}. \tag{4}$$

We begin our analysis of mixture models with the simplest case of exchangeable mixture models; more general models are considered in Section 4.3.

4.1 Exchangeable Mixture Models

A Bernoulli-mixture model will be termed *exchangeable* if the functions Q_i are all identical; in that case the random vector $\mathbf{Y}$ is exchangeable. Exchangeable Poisson-mixture models are defined analoguously; here $\Lambda_i(\cdot) = \Lambda$ for some function Λ with range $(0, \infty)$ and all $1 \leq i \leq m$.

We begin our analysis with exchangeable Bernoulli mixture models. It is convenient to introduce the rv $Q := Q_1(\mathbf{\Psi})$. We get for $\mathbf{y} = (y_1, \dots, y_m)'$ in $\{0,1\}^m$

$$P(\mathbf{Y} = \mathbf{y} \mid \mathbf{\Psi}) = Q_1(\mathbf{\Psi})^{\sum_{i=1}^m y_i} (1 - Q_1(\mathbf{\Psi}))^{m - \sum_{i=1}^m y_i} = P(\mathbf{Y} = \mathbf{y} \mid Q),$$

and, in particular, $P(Y_i = 1 \mid Q) = Q$. Denote by G the distribution function of Q. To calculate the unconditional distribution of $\mathbf{Y}$ or of the number of defaults M we integrate over the mixing distribution of Q to get

$$p(\mathbf{y}) = \int_0^1 q^{\sum_{i=1}^m y_i}(1-q)^{m-\sum_{i=1}^m y_i} dG(q)$$

$$P(M=k) = \binom{m}{k}\int_0^1 q^k(1-q)^{m-k} dG(q). \tag{5}$$

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

$$\pi_k = P(Y_1 = 1, \ldots, Y_k = 1) = E(E(Y_1 \cdots Y_k \mid Q)) = E(Q^k), \tag{6}$$

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 (see (1)) is always nonnegative. Any value of ρ_Y in $[0,1]$ can be obtained by an appropriate choice of the mixing distribution G. In particular, if $\rho_Y = \operatorname{var}(Q) = 0$ the rv Q has a degenerate distribution with all mass concentrated on the point π and the default indicators are independent. The case $\rho_Y = 1$ corresponds to a model where $\pi = \pi_2$ and the distribution of Q is concentrated on the points 0 and 1. In this model $Y_i = Q, \forall i$, and the default indicators are perfectly positively dependent (or comonotonic); see also Section 6.3 below.

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

- Beta mixing-distribution. Here $Q \sim \text{Beta}(a,b)$ with density $g(q) = \beta(a,b)^{-1}q^{a-1}(1-q)^{b-1}$, $a, b > 0$. This model is particularly tractable; see Example 4.3 below.
- Probit-normal mixing-distribution. Here $Q = \Phi(\Psi)$ for a rv Ψ following a normal distribution. This model can be viewed as a one-factor version of the CreditMetrics and KMV-type models; see Section 5.2.
- Logit-normal mixing-distribution. Here $Q = 1/(1+\exp(\Psi))$ for $\Psi \sim N(\mu, \sigma^2)$. 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.

Example 4.3 (Beta-Binomial Model). Of the choices above the most tractable is the beta distribution since it leads to simple closed form expressions for the integrals in (5) and (6). Using the well-known fact that $\beta(a+1,b)/\beta(a,b) = a/(a+b)$ we find that

$$\pi_k = \prod_{j=0}^{k-1} \frac{a+j}{a+b+j}, \qquad k = 1, 2, \ldots,$$

so that $\pi = a/(a+b)$, $\pi_2 = \pi(a+1)/(a+b+1)$ and $\rho_Y = (a+b+1)^{-1}$. Thus, if we fix any two of π, π_2 or ρ_Y, this determines a and b and the higher order joint default probabilities are automatic. Observe that $\pi_k/\pi_{k-1} = 1 - b/(a+b+k)$ increases with k, so that the conditional probability that a kth obligor defaults, given that some subgroup of $k-1$ obligors defaults increases with k. The number of defaults has a so-called beta-binomial distribution with probability function

$$P(M=k) = \binom{m}{k}\frac{\beta(a+k, b+m-k)}{\beta(a,b)}. \tag{7}$$

Calculations for the logit-normal, probit-normal and other models generally require numerical evaluation of integrals. If we fix any two of π, π_2 or ρ_Y in a logit-normal or probit-normal model, then this fixes the parameters μ and σ of the mixing distribution and higher order joint default probabilities are again automatic.

We now consider exchangeable Poisson mixture models. From (4) we get $P(\widetilde{M} = k \mid \Lambda) = e^{-m\Lambda}(m\Lambda)^k/k!$.

Example 4.4 (Gamma-Poisson Model). A very natural choice of mixing distribution for Λ is the gamma distribution $Ga(a,b)$ with density $g(\lambda) = b^a \lambda^{a-1} \exp(-b\lambda)/\Gamma(a)$, for constants $a, b > 0$. A direct computation shows that this leads to the following distribution for $\widetilde{M}$.

$$P(\widetilde{M} = k) = \left(1 - \frac{b}{m+b}\right)^k \left(\frac{b}{m+b}\right)^a \frac{\Gamma(k+a)}{\Gamma(k+1)\Gamma(a)}.$$

This is a negative binomial distribution with parameters a and $b/(m+b)$, which we denote $\widetilde{M} \sim Nb(a, b/(m+b))$. In the actuarial literature it is well-known that the negative binomial distribution can be obtained as Gamma-mixture of Poisson-distributions; see for instance Grandell (1997).

4.2 Loss distributions for large portfolios in exchangeable Bernoulli mixture models

We now show that in large exchangeable Bernoulli mixture models the quantiles of the credit loss distribution are essentially determined by the quantiles of the mixing distribution. Consider an exchangeable Bernoulli mixture model with mixing distribution given by the df G on $[0,1]$ and denote by $L^{(m)} = \sum_{i=1}^m E_i Y_i$ the credit-losses in a portfolio with m identical obligors. The positive rv E_i models the loss given that counterparty i defaults. We assume that $\{E_i\}_{i \in \mathbb{N}}$ is an iid sequence with mean μ_E and variance $\sigma_E^2 < \infty$, which is independent of the default indicators and the mixing variable.

Proposition 4.5. *Denote by $q_\alpha(Q)$ the α-quantile of the mixing distribution G of Q, i.e. $q_\alpha(Q) = \inf\{q,\ G(q) \geq \alpha\}$, and define $q_\alpha(L^{(m)})$ to be the α-quantile of the credit loss distribution in a portfolio with m obligors. Assume that the quantile function $\alpha \to q_\alpha(Q)$ is continuous in α, i.e. that*

$$G(q_\alpha(Q) + \delta) > \alpha \text{ for every } \delta > 0. \tag{8}$$

Then

$$\lim_{m \to \infty} \frac{1}{m} q_\alpha(L^{(m)}) = \mu_E\, q_\alpha(Q). \tag{9}$$

Remarks 4.6. 1) If the df G of the mixing distribution is strictly increasing, and in particular if G admits a density g which is a.e. positive on $[0,1]$ the condition (8) is satisfied for any $\alpha \in (0,1)$.
2) Consider two exchangeable Bernoulli mixture models with mixing distributions G_1 and G_2 and identical distribution of the loss given default. 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 4.5 implies that for large portfolios the tail of the credit loss distribution in the model with mixing distribution G_1 is heavier than the tail in the model with mixing distribution G_2. This has been observed in several simulation-studies including Gordy (2000); the above proposition explains this observation.
3) In the special case of the probit-normal mixing distribution, which corresponds to a one-factor version of the KMV-model, a similar limit result is obtained in (KMV-Corporation 1997).

Proof. Recall that conditional on $Q = q$ the losses $\{E_i Y_i\}_{i \in \mathbb{N}}$ form an iid sequence with mean $\mu_E q$. Hence we get from the law of large numbers or the central limit theorem for any $\lambda \geq 0$

$$\lim_{m \to \infty} P(L^{(m)} \leq \lambda m | Q = q) = \begin{cases} 1, \text{ if } \mu_E q < \lambda \\ 1/2 \text{ if } \mu_E q = \lambda \\ 0, \text{ if } \mu_E q > \lambda \end{cases}. \tag{10}$$

Now we get for any $\varepsilon > 0$

$$\begin{aligned}\limsup_{m\to\infty} P(L^{(m)} \le m(\mu_E\, q_\alpha(Q) - \varepsilon)) &= \limsup_{m\to\infty} \int_0^1 P\left(L^{(m)} \le m(\mu_E\, q_\alpha(Q) - \varepsilon) | Q = q\right) dG(q) \\ &\le \int_0^1 \limsup_{m\to\infty} P\left(L^{(m)} \le m(\mu_E\, q_\alpha(Q) - \varepsilon) | Q = q\right) dG(q) \\ &\le \int_0^1 1_{\{q \le\ q_\alpha(Q) - \varepsilon/\mu_E\}}\, dG(q) \\ &= G(q_\alpha(Q) - \varepsilon/\mu_E) < \alpha\,.\end{aligned}$$

Here the first inequality follows from Fatou's Lemma, the second inequality is a consequence of (10), and the last relation follows from the definition of the α-quantile. Similarly we have

$$\liminf_{m\to\infty} P(L^{(m)} \le m(\mu_E\, q_\alpha(Q) + \varepsilon)) \ge P(Q < q_\alpha(Q) + \varepsilon/\mu_E) > \alpha\,,$$

the last inequality being strict because of (8). Hence for m large enough we have $m(\mu_E q_\alpha(Q) - \varepsilon) \le q_\alpha(L^{(m)}) \le m(\mu_E q_\alpha(Q) + \varepsilon)$, which proves the claim. □

4.3 Mixture models with more general dependence structures

In this section we consider mixture models suitable for more heterogeneous portfolios where exchangeability cannot be assumed. One of the simplest ways to generalise the exchangeable mixture models is by constructing regression models. In these models deterministic covariates are allowed to influence the probability of default; the effective dimension of the mixing distribution is still one. These models may be particularly useful for statistical purposes. If an algorithm exists to fit an exchangeable model then this can be adapted easily to fit a regression model. We concentrate on Bernoulli regression models; Poisson regression model can be defined in an analogous manner.

Example 4.7 (Bernoulli regression model). Formally, a Bernoulli regression model is a one-factor Bernoulli mixture model with conditional default probabilities of the form $Q(\Psi, \mathbf{z_i})\,,\ 1 \le i \le m$.Here $\mathbf{z}_i \in \mathbb{R}^l$ is a vector of deterministic covariates and $Q : \mathbb{R} \times \mathbb{R}^l \to [0, 1]$ is strictly increasing in its first argument. There are many possibilities for this function; see for instance Joe (1997). Useful examples include

- $Q(\psi, \mathbf{z}_i) = F(\psi)^{\exp(-\boldsymbol{\beta}'\mathbf{z}_i)}$, where F is some continuous df on $\mathbb{R}$.
- $Q(\psi, \mathbf{z}_i) = F(\boldsymbol{\gamma}'\mathbf{z}_i\psi + \boldsymbol{\beta}'\mathbf{z}_i)$ where $\boldsymbol{\gamma}'\mathbf{z}_i > 0$.

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

The regression model may not be flexible enough to model the true heterogeneity we observe in a large portfolio. We can generalise the model by increasing the dimension of the mixing distribution through the assumption of more common stochastic factors.

Example 4.8. We now introduce a Bernoulli mixture model with factor structure where the factors have a Gaussian dependence structure; we will see in Section 5.2 below that this model is closely related to the models proposed by KMV and CreditMetrics. Assume that

$\mathbf{\Psi} \sim N_p(0, I)$ and let $\mathbf{w}_i = (w_{i,1}, \dots, w_{i,p})'$ be a vector of weights for obligor i. Let $\mathbf{c}$ be an m-dimensional vector of constants. We put

$$Q_i(\boldsymbol{\psi}) = \Phi\left(c_i - \sum_{j=1}^{p} w_{i,j}\psi_j\right),$$

where Φ is the standard normal distribution function. Thus each individual default probability has a probit-normal distribution where $\Phi^{-1}(Q_i) \sim N(c_i, \mathbf{w}_i'\mathbf{w}_i)$.

We now consider an industry-example of a Poisson-mixture model with factor structure.

Example 4.9 (CreditRisk$^+$). CreditRisk$^+$is essentially a Poisson mixture model with factor structure. For the mixing variables in Definition 4.2 it is assumed that

$$\Lambda_i(\mathbf{\Psi}) = \mathbf{w}_i'\mathbf{\Psi},\ \forall i, \tag{11}$$

where $\mathbf{\Psi} = (\Psi_1, \dots, \Psi_p)'$ is a random vector of independent gamma distributed risk factors with $p < m$ and $\mathbf{w}_i = (w_{i,1}, \dots w_{i,p})'$ is now a vector of non-negative factor weights for obligor i. To calibrate the model obligors are divided into rating classes, for which it is assumed that individual default rates are constant so that $E(\Lambda_i(\mathbf{\Psi})) = \overline{q}_{g(i)}$ where $\overline{q}_{g(i)}$ represents an estimate of the default rate for all obligors in the group $g(i)$ to which obligor i belongs.

We now show that the rv $\widetilde{M} := \sum_{i=1}^{m} \widetilde{Y}_i$ is equal in distribution to a sum of independent negative binomial random variables, which facilitates the application of the model. Suppose that $\Psi_j \sim Ga(a_j, b_j)$, $j = 1, \dots, p$. Conditional on $\mathbf{\Psi}$ the rv $\widetilde{M}$ is then Poisson-distributed with parameter

$$\sum_{i=1}^{m}\sum_{j=1}^{p} w_{ij}\Psi_j = \sum_{j=1}^{p}\Psi_j\left(\sum_{i=1}^{m} w_{ij}\right).$$

Now consider p random-variables $\widetilde{M}_j$, $1 \le j \le p$, which conditional on $\mathbf{\Psi}$ are independent Poisson$(\Psi_j\,(\sum_{i=1}^{m} w_{ij}))$-distributed. Clearly, by the independence of the Ψ_j the $\widetilde{M}_j$ are independent and $\widetilde{M} \stackrel{\mathrm{d}}{=} \sum_{j=1}^{p} \widetilde{M}_j$. Moreover, $\widetilde{M}_j$ has a negative binomial distribution since $\Psi_j \sum_{i=1}^{m} w_{ij} \sim Ga(a_j, b_j / \sum_{i=1}^{m} w_{ij})$; cf Example 4.4.

4.4 Multinomial mixture models

For reasons of notational simplicity and ease of understanding we have so far dealt exclusively with Bernoulli mixture models for two states (default and non-default). We now show briefly that these ideas can be generalised to multinomial mixture models, which allow us to model defaults and rating transitions over the period of interest. To define these models we introduce the notation $\mathcal{S}_n = \{(q_0, q_1, \dots, q_n)' \in \mathbb{R}^{n+1} : 0 \le q_j \le 1, \forall j; \sum_{j=0}^{n} q_j = 1\}$ for the n-dimensional simplex, which is the support of our mixing distribution in the general multi-state model. The definition follows the pattern of Definition 4.1 and Definition 4.2.

Definition 4.10 (Multinomial Mixture Model). Given some $p < m$ and a p-dimensional random vector $\mathbf{\Psi} = (\Psi_1, \dots, \Psi_p)$ then for $n \in \mathbb{N}$ the random state vector $\mathbf{S} = (S_1, \dots, S_m)'$ follows an n-dimensional multinomial mixture model with factor vector $\mathbf{\Psi}$, if there are functions $\mathbf{Q}_i : \mathbb{R}^p \to \mathcal{S}_n$, $1 \le i \le m$, such that conditional on $\mathbf{\Psi}$ the state vector $\mathbf{S}$ is a vector of independent multinomial random variables with

$$P(S_i = j \mid \mathbf{\Psi}) = [\mathbf{Q}_i(\mathbf{\Psi})]_j\,, \quad i = 1, \dots, m, j = 0, 1, \dots, n,$$

where $[.]_k$ denotes the kth element of a vector and indexing starts at zero.

Remark 4.11. A 1-dimensional multinomial mixture model corresponds to a two-state model, which is equivalent to a Bernoulli mixture model.

For the mixing distribution on the simplex there are again many possibilities, but the most analytically tractible example is the Dirichlet distribution, a multivariate distribution whose margins are beta. The following exchangeable mixture model generalises the beta-binomial model of Example 4.3.

Example 4.12 (Dirichlet-Multinomial Model). Consider a homogeneous portfolio of m non-defaulted obligors. For $j = 0, 1, \ldots, n$ let K_j denote the number of obligors who have migrated to state j bt time T. Assume an exchangeable model with $\mathbf{Q}_1(\mathbf{\Psi}) = \cdots = \mathbf{Q}_m(\mathbf{\Psi}) =: \mathbf{Q}$ and denote the df of $\mathbf{Q}$ by $G(\mathbf{q}) := G(q_0, \ldots, q_n)$. For a general mixing distribution we have

$$P(K_0 = k_0, \ldots, K_n = k_n) = \binom{m}{k_0 \ldots k_n} \int_{\mathbf{q} \in \mathcal{S}_n} q_0^{k_0} \cdots q_n^{k_n} dG(\mathbf{q}), \quad \sum_{j=0}^{n} k_j = m.$$

The density of the Dirichlet is given by

$$g(\mathbf{q}) = \Gamma\left(\sum_{j=0}^{n} a_j\right) \prod_{j=0}^{n} \frac{q_j^{a_j - 1}}{\Gamma(a_j)}, \quad \mathbf{q} \in \mathcal{S}_n, \quad a_j > 0, j = 0, 1, \ldots, n,$$

and simple calculations show that in this case

$$P(K_0 = k_0, \ldots, K_n = k_n) = \binom{m}{k_0 \ldots k_n} \frac{\Gamma\left(\sum_{j=0}^{n} a_j\right)}{\Gamma\left(\sum_{j=0}^{n} (a_j + k_j)\right)} \prod_{j=0}^{n} \frac{\Gamma(a_j + k_j)}{\Gamma(a_j)}.$$

It may be verified that when $n = 1$ this model reduces to the beta-binomial distribution in (7).

5 Relationship between the model types

At a first glance latent variable models and Bernoulli mixture models appear to be very different types of models. However, as has already been observed by Gordy (2000), these differences are often related more to presentation and interpretation than to mathematical substance. Gordy showed that the CreditMetrics and KMV-type models can be written as Bernoulli mixture models if the asset returns have a factor structure. In this section we generalize this observation and provide a fairly general result linking latent variable models and mixture models. This result and many of the other ideas we present extend quite naturally to multinomial mixture models, but we again concentrate on the Bernoulli case for clarity of presentation and restrict ourselves to a couple of remarks concerning the extension to the multinomial case.

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; see for instance the excellent discussion in Gordy (2000). 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 latent variable models also to mixture models and vice versa. For instance, we will see in Section 6 below that the distribution of the number of defaults in a CreditMetrics-type latent variable model can be determined much faster if we simulate Bernoulli random variables from the corresponding mixture model.

5.1 A general result

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

Definition 5.1. A latent-variable-vector $\mathbf{X}$ has a p-dimensional conditional independence structure with conditioning variable $\mathbf{\Psi}$, if there is some $p < m$ and a p-dimensional random vector $\mathbf{\Psi} = (\Psi_1, \ldots, \Psi_p)$ such that conditional on $\mathbf{\Psi}$ the rv's $(X_i)_{1 \le i \le m}$ are independent.

Interesting examples of latent variable models with a p-dimensional conditional independence structure are provided by multivariate normal mixture models with factor structure. These models are discussed in more detail below.

Proposition 5.2. *Consider an m-dimensional latent variable vector $\mathbf{X}$ with continuous marginal distributions and a p-dimensional ($p < m$) random vector $\mathbf{\Psi}$. Then the following are equivalent.*

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

Proof. Suppose that (i) holds. We have for $y \in \{0,1\}^m$

$$\begin{aligned} P(Y_1 = & y_1, \ldots, Y_m = y_m \mid \mathbf{\Psi}) = \\ &= P((-1)^{1-y_1} X_1 < (-1)^{1-y_1} D_0^1, \ldots, (-1)^{1-y_m} X_m < (-1)^{1-y_m} D_0^m \mid \mathbf{\Psi}) \\ &= \prod_{i=1}^m P((-1)^{1-y_i} X_i < (-1)^{1-y_i} D_0^i \mid \mathbf{\Psi}). \end{aligned}$$

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

Remark 5.3. This result extends very easily to the multinomial case. (ii) may be replaced by the equivalent condition that for any sequence of cut-off levels $-\infty = D_{-1}^i < D_0^i < \cdots < D_n^i = \infty$, $1 \le i \le m$, the state indicator vectors $\mathbf{S}_i$ given by $S_i = j \iff D_{j-1}^i < X_i \le D_j^i$, $j = 0, \ldots, n$, follow an n-dimensional multinomial mixture model with factor vector $\mathbf{\Psi}$. The probability of being in state j is given by $[\mathbf{Q}_i(\mathbf{\Psi})]_j = P(D_{j-1}^i < X_i \le D_j^i \mid \mathbf{\Psi})$.

5.2 Normal mean-variance mixtures with factor structure

This class of models is constructed from an m-dimensional random vector $\mathbf{Z} \sim N_m(\mathbf{0}, \Sigma)$ with a p-dimensional linear factor structure for some $p < m$. For more details of the linear factor model see Appendix B. Basically we assume that the components of $\mathbf{Z}$ can be written as

$$Z_i = \sum_{j=1}^p a_{i,j} \Theta_j + \sigma_i \varepsilon_i, \tag{12}$$

for a p-dimensional random vector $\mathbf{\Theta} \sim N_p(\mathbf{0}, \Omega)$ and independent standard normally distributed rv's $\varepsilon_1 \ldots, \varepsilon_m$, which are also independent of $\mathbf{\Theta}$. Obviously $\mathbf{Z}$ has p-dimensional conditional independence structure when we condition on $\mathbf{\Theta}$.

Consider now a random variable W which is independent of $\mathbf{Z}$ and functions $\mu_i : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to (0, \infty)$. Assuming that $\mathbf{Z}$ is of the form (12), define the latent variables $\mathbf{X}$ by

$$X_i := \mu_i(W) + g(W) Z_i, \quad 1 \le i \le m. \tag{13}$$

Then $\mathbf{X}$ has a $(p+1)$-dimensional conditional independence structure. Define the $(p+1)$-dimensional random vector $\mathbf{\Psi}$ by $\mathbf{\Psi} := (\Theta_1, \ldots, \Theta_p, W)'$. Conditional on $\mathbf{\Psi}$ the rv's X_i are obviously independent normally distributed with mean $\mu_i(W) + g(W)\sum_{j=1}^{p} a_{ij}\Theta_j$ and variance $(g(W)\sigma_i)^2$. The equivalent Bernoulli mixture model is now easy to compute. Given thresholds $(D_0^i)_{1\le i\le m}$ we get that

$$Q_i(\mathbf{\Psi}) = P(X_i < D_0^i \mid \mathbf{\Psi}) = \Phi\left(\frac{D_0^i - \mu_i(W) - g(W)\sum_{j=1}^{p} a_{ij}\Theta_j}{g(W)\sigma_i}\right). \tag{14}$$

Example 5.4 (Gaussian latent variables). In the special case where no mixing takes place such that $\mathbf{X} = \mathbf{Z}$, the equivalent mixture model has the form given in Example 4.8. This is the Bernoulli mixture model corresponding to the CreditMetrics/KMV-type models; it is also derived in Gordy (2000). Of course any latent variable model where the copula is that of a multivariate normal distribution with p-factor structure will lead to this type of Bernoulli-mixture model.

Example 5.5 (Student t latent variables). If we use construction (13) and take

$$\mu_i \equiv 0\,,\ i = 1, \ldots, m\,, \quad g(w) = \sqrt{\frac{\nu}{w}}\,,\ \text{and}\ \ W \sim \chi^2(\nu), \tag{15}$$

then $\mathbf{X}$ has an m-dimensional t distribution with ν degrees of freedom, mean $\mathbf{0}$ and, for $\nu > 2$, covariance matrix $\frac{\nu}{\nu-2}\Sigma$. This is usually denoted by $t_m(\nu, \mathbf{0}, \Sigma)$. Note that the correlation-matrix of $\mathbf{X}$ equals the correlation matrix of $\mathbf{Z}$ and $\mathbf{X}$ inherits the linear factor structure of $\mathbf{Z}$. This property holds more generally for all normal variance mixtures, where the mean does not depend on the mixing variable W. Normal variance mixtures are important examples of elliptical distributions; see for instance Embrechts, McNeil, and Straumann (2001) for more on elliptical distributions in risk management. The default indicators in the t-model follow a $(p+1)$-factor Bernoulli mixture model with conditional default probability given by

$$Q_i(\mathbf{\Psi}) = \Phi\left(\sigma_i^{-1}\left(D_0^i\sqrt{W/\nu} - \sum_{j=1}^{p} a_{ij}\Theta_j\right)\right). \tag{16}$$

Example 5.6 (Generalized hyperbolic latent variables). This is an example of a full mean-variance mixture which is no longer an elliptical distribution. To obtain a generalized hyperbolic distribution we assume that the mixing variable W follows a normal inverse Gaussian distribution and take $\mu_i(W) = \beta_i W^2$ for constants β_i and $g(W) = W$. The generalized hyperbolic distribution has been advocated by Eberlein and Keller (1995) as a model for stock returns.

5.3 Relationship between the model types: the exchangeable case

In the special case of exchangeable default indicators the connection between the model types is related to well-known results of De Finetti and Hausdorff. We refer to Feller (1971) for important background material to this section.

De Finetti's theorem concerns *infinite* sequences of exchangeable Bernoulli random variables; extensions to random variables which can take an arbitrary but finite number of values, such as the multinomial distribution, are also available. The infinite sequence $Y_1, Y_2, \ldots$ is said to be exchangeable if the random vectors $(Y_1, \ldots, Y_k)$ are exchangeable for all $k \in \mathbb{N}$. De Finetti's theorem essentially shows that infinite exchangeable Bernoulli sequences can be modelled by an infinite exchangeable Bernoulli mixture.

Theorem 5.7 (De Finetti). *For any infinite sequence $Y_1, Y_2, \ldots$ of exchangeable Bernoulli random variables there is a probability distribution G on $[0,1]$ such that for all $k \le m \in \mathbb{N}$*

$$P(Y_1 = 1, \ldots, Y_k = 1, Y_{k+1} = 0, \ldots, Y_m = 0) = \int_0^1 q^k (1-q)^{m-k} dG(q). \tag{17}$$

We can use this result to show that certain latent variable models induced by an exchangeable copula, which are not immediately seen to have a conditional independence structure, can be represented as exchangeable Bernoulli mixtures. We consider latent variable models defined using Archimedean copulas; see for instance Nelsen (1999). A k-dimensional Archimedean copula is the distribution function of an exchangeable uniform random vector and has the form

$$C_{1,\ldots,k}(u_1,\ldots,u_k) = \phi^{-1}\left(\phi(u_1) + \cdots + \phi(u_k)\right), \tag{18}$$

where $\phi : [0,1] \mapsto [0,\infty]$ is a continuous, strictly decreasing function, known as the copula *generator* which satisfies $\phi(0) = \infty$ and $\phi(1) = 0$; ϕ^{-1} is the generator inverse. A theorem of Kimberling (1974) (see also Schweizer and Sklar (1983)) shows that a necessary and sufficient condition for (18) to define a proper copula for all k is that ϕ^{-1} is a *completely monotonic* function on $[0,\infty)$, i.e. $(-1)^k \frac{d^k}{dt^k}\phi^{-1}(t) \geq 0, \quad k \in \mathbb{N}$.

Observe that for Archimedian copulas $C_{1,\ldots,k}$ is the k-dimensional marginal distribution of $C_{1,\ldots,k+1}$. This allows us to construct an *infinite* sequence of exchangeable uniform random variables U_n, $n \in \mathbb{N}$ whose k-dimensional margin is equal to $C_{1,\ldots,k}$ for all $k \in \mathbb{N}$. The existence of such a sequence sequence is the key to the following result.

Proposition 5.8. *Consider a latent variable model* $\left(X_i, (D^i_j)_{-1\leq j\leq n}\right)_{1\leq i\leq m}$ *and suppose that* $\mathbf{X}$ *has an exchangeable Archimedean copula* (18) *with completely monotonic generator. Assume* $P(X_i \leq D^i_0) = \pi, \forall i$. *Then the default indicators* $Y_i = 1_{\{X_i \leq D^i_0\}}$ *follow an exchangeable Bernoulli mixture model.*

Proof. The existence of an infinite sequence U_n, $n \in \mathbb{N}$ of rv's with k-dimensional margin equal to $C_{1,\ldots,k}$ is equivalent to the existence of a probability measure μ_∞ on $[0,1]^\infty$ whose k-dimensional marginal equals $C_{1,\ldots,k}$ for all $k \in \mathbb{N}$. Let μ_1 be the uniform distribution on $[0,1]$. At the $(k+1)$th step use the conditional distribution $C_{k+1|1,\ldots,k}(U_{k+1} \mid U_1,\ldots U_k)$ to define a transition kernel K_k from $[0,1]^k$ to $[0,1]$ and define the measure μ_{k+1} on $[0,1]^{k+1}$ by $\mu_{k+1} := \mu_k \otimes K_k$; obviously $\mu_k = C_{1,\ldots,k}$ for all k. The existence of μ_∞ now follows from the Ionescu-Tulcea theorem (see for instance Theorem 3, Chapter 22 in Fristedt and Gray (1997)).

If we define $\widetilde{Y}_i = 1_{\{U_i \leq \pi\}}, i = 1, 2, \ldots$ we have an infinite exchangeable Bernoulli sequence. Observe that $\left(X_i, D^i_0\right)_{1\leq i\leq m}$ and $(U_i, \pi)_{1\leq i\leq m}$ are two equivalent latent variable modes by Proposition 3.3. Therefore $(Y_1,\ldots,Y_m)$ and $(\widetilde{Y}_1,\ldots,\widetilde{Y}_m)$ have the same multivariate Bernoulli distribution. But $\widetilde{Y}_1,\ldots,\widetilde{Y}_m$ form the first m terms of an infinite exchangeable Bernoulli sequence and thus De Finetti shows the existence of a random variable Q with support in $[0,1]$ so that

$$\pi_k = P(Y_1 = 1,\ldots,Y_k = 1) = \phi^{-1}\left(k\phi(\pi)\right) = E\left(Q^k\right), \quad k = 1, 2, \ldots. \tag{19}$$

□

Remarks 5.9. 1) This result also extends to the multinomial case. If we define state vectors $\widetilde{S}_i$ by dividing the range of U_i into more than two intervals then we can construct infinite exchangeable multinomial sequences and appeal to variants of De Finetti for the existence of mixing distributions.

2) With Archimedean copulas of the form (18) it is very easy to calculate higher order joint default probabilities $\pi_k = \phi^{-1}(k\phi(\pi))$, which are the moments of the equivalent mixing variable Q. From these it is theoretically possible to calculate, or at least approximate, the distribution function G of Q. To do this we could make use of an approximation result found for instance in Feller (1971), which shows how a distribution G on $[0,1]$ is determined

by iterated differences between its moments. Define $\Delta^1 \pi_k := \pi_{k+1} - \pi_k$ and recursively $\Delta^j \pi_k = \Delta^1 \Delta^{j-1} \pi_k, j = 2, 3 \ldots$. Then the distribution with mass

$$p_k^n := \binom{n}{k} (-1)^{n-k} \Delta^{n-k} \pi_k$$

on the points k/n, $1 \leq k \leq n$, converges to G as $n \to \infty$.

Example 5.10. There are many possibilities for generating Archimedean copulas (Nelsen 1999). If we take the generator $\phi_\theta(t) = t^{-\theta} - 1$ we get Clayton's copula family; it may be verified that the generator inverse is a completely monotonic function.

De Finetti's theorem shows that any exchangeable model for $\mathbf{Y}$ which can be extended to arbitrary portfolio size m has a representation as exchangeable Bernoulli-mixture model. In order to further illustrate the difference between finite and infinite exchangeable models we now adapt an example of Feller (1971) and construct a latent variable model yielding a finite exchangeable Bernoulli sequence which does not follow an exchangeable Bernoulli mixture model. Clearly, the example is of limited practical relevance, but it serves to illustrate that for a fixed portfolio size m the class of m-dimensional exchangeable latent variable models is larger than the class of exchangeable Bernoulli mixtures. To do this we use a theorem of Hausdorff.

Theorem 5.11 (Hausdorff). *A sequence $\pi_1, \pi_2, \ldots$ represents the moments of a probability distribution G with support in $[0, 1]$ if and only if it is completely monotonic, i.e.*

$$(-1)^j \Delta^j \pi_k \geq 0, \quad \pi_0 = 1. \tag{20}$$

Example 5.12. Consider three oil companies ($m = 3$), and assume that each of these companies owns drilling rights for a particular area in a new oilfield. There are two exploitable petroleum reservoirs in the field, although these have yet to be found. Assume that all 6 different possibilities of allocating these reservoirs among the claims (two on claim 1, two sources on claim 2, and so on) have equal probability 1/6. Let Z_i denote the number of reservoirs in the possession of company i; clearly $Z_i \in \{0, 1, 2\}$. Now take three independent rv's $W_i \sim U(0, 1)$ and let $X_i = W_i + Z_i$ be the asset value of company i. Set the liabilities to be $D_i^0 = 1$ for all companies. $(X_i, D_i^0)_{1 \leq i \leq 3}$ defines a latent variable model yielding a trivariate exchangeable Bernoulli vector (Y_1, Y_2, Y_3), where $Y_i = 1_{\{X_i \leq D_i^0\}} = 1_{\{Z_i = 0\}}$, i.e. a company defaults if and only if there is no oil on its claim. It is easily calculated that $\pi_1 = 1/2, \quad \pi_2 = 1/6, \quad \pi_3 = 0$. If this were the beginning of a completely monotonic sequence we would have $\pi_4 = \pi_5 = \cdots = 0$. But in this case $\Delta^4 \pi_1 < 0$ contradicting (20), so these cannot be the moments of a distribution with support in $[0, 1]$.

Note however, that $\mathbf{X}$ has a two-dimensional conditional independence structure with conditioning variable $\mathbf{\Psi} = (Z_1, Z_2)$ and can therefore be written as two-factor Bernoulli-mixture model.

6 Sensitivity of Losses to Dependence Specification

6.1 Fixing default-probability and asset correlation

In the CreditMetrics and KMV approaches the correlations of the latent variables $\mathbf{X}$, representing changes in asset values, are modelled by a linear factor model (see Appendix B). Suppose that a factor model has been set up and that factor weights have been determined. This fixes the correlation matrix R of $\mathbf{X}$, but not the multivariate distribution of $\mathbf{X}$ and, in particular, not its copula. Here we are interested in the effect of different distributional assumptions for $\mathbf{X}$ on the distribution of the number of defaults M, assuming that the asset correlation matrix R has been fixed. We are particularly interested in the sensitivity of the credit loss distribution with respect to the assumption of multivariate normality, since popular models such as CreditMetrics and KMV assume a Gaussian copula; cf Section 3.

We compare two models, a model with Gaussian latent variables and a model where latent variables follow a t distribution. For simplicity we conduct this investigation in the setting of an exchangeable one-factor model, i.e. we model Z in (12) as

$$Z_i = \sqrt{\rho}\Theta + \sqrt{1-\rho}\varepsilon_i, \quad i = 1, \ldots, m, \quad \rho \geq 0\,.$$

In the normal case we put $X_i = Z_i$, $i = 1, \ldots, m$; in the t case we put $X_i = \sqrt{\nu/W} Z_i$ for a rv $W \sim \chi^2(\nu)$ independent of $\mathbf{Z}$ (see Example 5.5). In both cases we choose cut-off levels so that $P(Y_i = 1) = \pi$, $\forall i$.

Note that for both models the asset correlation matrix R is given by aa equicorrelation matrix with off-diagonal element ρ. In the first case the copula of $\mathbf{X}$ is the exchangeable Gaussian copula C_ρ^{Ga} which has a single parameter ρ (see Appendix A for more details). In the second $\mathbf{X}$ has an exchangeable t copula, denoted by $C_{\nu,\rho}^{\mathrm{t}}$, which has an extra parameter ν. Higher order joint default probabilities take the form given in (3), and we expect more defaults in the t model, due to the tail dependence of the t copula.

We conduct a simulation study where we vary the portfolio size m, the individual default probabilities π, the correlation of the latent variables ρ and the degrees of freedom ν of the t copula. To perform the simulation we use the equivalent Bernoulli mixture model representations, since this is more efficient than simulating the latent variables directly; the conditional default probabilities are given in (14).

We define 3 groups of decreasing credit quality, which we label A, B and C. In Group A we set $\pi = 0.06\%$ and $\rho = 2.58\%$; in Group B we set $\pi = 0.50\%$ and $\rho = 3.80\%$; in Group C $\pi = 7.50\%$ and $\rho = 9.21\%$. These do not correspond exactly to the A, B and C rating categories used by any of the well-known rating agencies, but they are nonetheless realistic values for Gaussian latent variable models for real obligors and were chosen after discussions with a Swiss bank. Similar simulation studies are also reported in Frey, McNeil, and Nyfeler (2001).

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

Table 1: Results of Simulation study. Estimated 95th and 99th percentiles of the distribution of M, the number of defaulting obligors, in an exchangeable model. See text for the values of π and ρ corresponding to the 3 groups A, B and C. Note that the quantiles are approximately proportional to the size of the portfolio; this shows that the asymptotic result of Proposition 4.5 is useful even for relatively small portfolios.

In all simulations we generate 100000 realisations of M. Of course $E(M) = mE(Q) = m\pi$ in all cases, and it is easily confirmed that the empirical average number of defaults is always very close to $m\pi$. Of greater interest are high quantiles of the distribution of M which give a better indication of the extreme risk in the model. We denote the empirically estimated 95% and 99% quantiles of the distribution of the number of defaults M by $\widehat{q_{0.95}}(M)$ and $\widehat{q_{0.99}}(M)$ respectively and tabulate them in Table 1. In Figure 1 we plot the ratio of estimated quantiles for a Student t model with 10 degrees of freedom and a Gaussian model in the case of Group B and a portfolio of size 10000.

Clearly ν has a massive influence on these risk measures, particularly for groups of poorer credit quality (B and C) and larger portfolio sizes. If we only specify the latent variable correlation ρ and do not fix the degrees of freedom ν then our inference concerning extreme risk is subject to huge model risk. This simulation study indicates that an attenpt to calibrate latent variable models based on marginal default probabilities and assumptions about latent variable correlations alone is not advisable.

The cause of these differences is clearly seen in Figure 2 where the densities of $\Phi^{-1}(Q)$ are shown for Group B in all four cases (Gaussian, $\nu = 50, 10, 4$). For all of these distributions $E(Q) = \pi = 0.005$. The heavier the right tail of the distribution of $\Phi^{-1}(Q)$ (or equivalently of Q) the heavier the tail of the distribution of M in sufficiently large portfolios.

6.2 Fixing default-probability and default correlation

In this Section we look at exchangeable Bernoulli mixture models where the default probability π and the default correlation ρ_Y, or equivalently the joint default probability π_2, are assumed to be known and fixed. We define two groups of obligors, B and C, with equal π and ρ_Y. The values of the parameters π, ρ_Y and π_2 are given in Table 2. The π_2 (or ρ_Y) values are the values implied by the latent variable correlation values ρ of the previous section in the case where latent variables are multivariate Gaussian; in other words $\pi_2 = C_\rho^{\mathrm{Ga}}(\pi, \pi)$.

Group	π	π_2	ρ_Y
B	0.005	0.000034	0.0018
C	0.075	0.007650	0.0292

Table 2: Values of π, π_2 and ρ_Y for groups B and C.

We compare the Gaussian and Student t latent variable models as well as the beta and logit-normal mixture models. The beta is easily calibrated to given π and π_2 values using the explicit formulae from Example 4.3, whereas a combination of numerical integration and root-finding is needed to accurately calibrate the other models. In the t model there is the added subtlety that for given ν it may be impossible to find a latent variable correlation value ρ which yields the desired default correlation. Unlike the Gaussian copula, the t copula does not give independence when $\rho = 0$ and we have $C^t_{\nu,0}(\pi, \pi) > \pi^2$. If default correlation is weak, as it often is in groups of better credit quality, we may have to choose ν high in order to calibrate the t model. In our study we choose $\nu = 100$ when we work with Group B and $\nu = 20$ when we work with Group C.

As before we simulate 100000 realisations of M for portfolio sizes of 1000 and 10000 drawn from groups B and C. We tabulate empirical estimates of the 95% and 99% quantiles of the distribution of M in Table 3. Clearly the quantile estimates now differ much less and we have to go to the 99% level to see small differences emerging. The Gaussian model is slightly riskier (in the quantile sense) than both the t model and the Beta mixture model; the logit-normal mixture model is slightly riskier than the Gaussian model.

These differences can again be explained by differences in the tail of the underlying mixture distribution of Q, cf Remark 4.6, although this time the differences are more difficult to see in a density plot. In Figure 3 we plot the tail function on a logarithmic y-scale in the Gaussian, beta and logit-normal cases. (The t case is more difficult to plot as the df of the mixture distribution is not easily calculated.) Although the distributions are very similar up to the 99th percentile, they begin to diverge in the way we expect beyond that point. If we were interested in higher percentiles of the loss distribution these differences would be meaningful.

m	Group	$\widehat{q_{0.95}}(M)$				$\widehat{q_{0.99}}(M)$			
		Gauss	t	Beta	Logit-N	Gauss	t	Beta	Logit-N
1000	B	12	12	12	12	17	17	17	18
1000	C	163	163	163	163	222	221	216	231
10000	B	109	109	109	108	155	154	148	158
10000	C	1612	1617	1615	1623	2214	2181	2141	2294

Table 3: Results of Simulation study. Estimated 95th and 99th percentiles of the distribution of M, the number of defaulting obligors, in various exchangeable Bernoulli mixture model with first two moments of mixing distribution fixed.

6.3 Asymptotically worst exchangeable Bernoulli mixtures

Recall from Section 4.2 and in particular from Remark 4.6 that in a large exchangeable Bernoulli mixture the tail of the loss distribution is essentially determined by the tail of the mixing distribution. This motivates the following definition

Definition 6.1. Consider two mixing-distributions μ and ν with df G_μ and G_ν on $[0,1]$. Then μ is called *asymptotically worse* than ν (in tail order), if the tail of G_μ dominates the tail of G_ν, i.e. if there is some $\delta > 0$ such that $G_\mu(q) \leq G_\nu(q)$ for all $q \in [1-\delta, 1]$ and moreover $\int_{1-\delta}^{1}(G_\nu(q) - \mathcal{G}_\mu(q))dq > 0$.

We can derive the simple forms of the worst-case mixing distributions when the first moment π or the first two moments π and π_2 (or π and $\rho_Y = (\pi_2 - \pi^2)/(\pi - \pi^2)$) are fixed. Of course the resulting worst-case mixing distributions are not to be interpreted as realistic models for modelling dependency between real world credit losses; rather their purpose is to illustrate how much model risk remains if we only specify default probability and default correlation.

The following lemma shows that finding asymptotically worst distributions amounts to maximising the probability that the mixing variable Q equals 1; that is maximising the probability that the entire exchangeable group defaults.

Lemma 6.2. *Given two mixing distributions μ and ν. Suppose that $\mu(\{1\}) > \nu(\{1\})$. Then μ is asymptotically worse than ν.*

Proof. We have

$$\lim_{q\to1} G_\mu(q) = G_\mu(1-) = 1 - \mu(\{1\}) > 1 - \nu(\{1\}) = \lim_{q\to1} G_\nu(q)\,,$$

and hence $G_\mu(q) > G_\nu(q)$ for q close to 1. □

1.) Asymptotically worst distributions with given default-probability π
In light of Lemma 6.2 we have to find a distribution μ maximizing $\mu(\{1\})$ under the constraint that $E^\mu(Q) := \int_0^1 q dG_\mu(q) = \pi$. Obviously for any such distribution $\mu(\{1\}) \leq E^\mu(Q) = \pi$. Hence the two-point distribution μ_1^* with $\mu_1^* = (1-\pi)\delta_0 + \pi\delta_1$ and δ_x the Dirac-measure in x is the corresponding worst-case distribution.

2.) Asymptotically worst distributions with given π and ρ_Y
This case is more interesting. By Jensen's inequality we must have $\pi_2 \in [\pi^2, \pi]$; the boundary cases $\pi_2 = \pi^2$ and $\pi_2 = \pi$ correspond to the mixing distributions $\mu = \delta_\pi$ (independent defaults) and $\mu = \mu_1^*$ (comonotonic defaults, $\rho_Y = 1$), respectively.

Proposition 6.3. *Fix $\pi \in (0,1)$ and $\pi_2 \in [\pi^2, \pi]$. The asymptotically worst mixing distribution with first two moments equal to π and π_2 is the two-point distribution μ_2^* with*

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

Related results on bounds on stop-loss premia for distributions on finite intervals with given first and second moment can be found in the actuarial literature; see for instance Chapter 5.5 of Goovaerts, De Vylder, and Haezendonck (1984).

Proof. Elementary calculations show that μ_2^* satisfies the moment constraints. Optimality of μ_2^* can be shown very elegantly by means of the *one-sided Chebyshev inequality.*[1] The one-sided Chebyshew inequality states that for any rv X on the real line on the real line with mean $E(X)$ and finite variance $\mathrm{var}(X)$ and any $\delta > 0$

$$P(X > E(X) + \delta) \le \frac{\mathrm{var}(X)}{\delta^2 + \mathrm{var}(X)}; \tag{22}$$

see Marshall and Olkin (1960) or Bertsimas and Popescu (2000) for a proof. Using (22) we get for any mixing distribution μ satisfying our moment constraints

$$P^\mu(Q=1) = P^\mu(Q > \pi + (1-\pi)) \le \frac{\pi_2 - \pi^2}{(1-\pi)^2 + \pi_2 - \pi^2} = p^*,$$

which shows that μ_2^* is in fact optimal. □

Our previous analysis has underlined the importance of the moments of the mixing distribution or equivalently the joint default probabilities $(\pi_k)_{k=1,2,\ldots}$ for the properties of the tail of the credit loss distribution. Moreover, in certain cases, such as the mixing models corresponding to the Archimedian copulas, the π_k are easily available whereas we do not have an explicit formula for the mixing distribution; see Remark 5.9 (part 2). Hence it is natural to ask, if we can use the sequence π_k to decide if one distribution is asymptotically worse than another and vice versa. The following proposition and its corollary give a partial answer to this question.

Proposition 6.4. *Given two distributions μ amd ν with df G_μ resp G_ν on $[0,1]$. Suppose that μ is asymptotically worse than ν in the sense of Definition 6.1. Then μ is asymptotically worse than ν in moment order, i.e. there is some k_0 such that*

$$\pi_k^\mu := \int_0^1 q^k dG_\mu(q) > \pi_k^\nu := \int_0^1 q^k dG_\nu(q) \text{ for all } k > k_0\,. \tag{23}$$

It follows from Lemma 6.2 and the previous proposition that if two mixing-distributions μ and ν satisfy $\mu(\{1\}) > \nu(\{1\})$ then μ is asymptotically worse than ν in moment order.

A partial converse of Proposition 6.4 is the following.

Corollary 6.5. *Consider two mixing-distributions μ and ν with df G_μ and G_ν on $[0,1]$. Suppose that the two df's are ordered in the sense of Definition 6.1, i.e. either μ is asymptotically worse than ν in tail order or vice versa. If μ is asymptotically worse than ν in moment order it must be the case that μ is asymptotically worse than ν in tail order.*

Essentially the corollary says that if we can assume that two mixing distributions μ and ν are ordered in tail order we can use the sequence of moments to check if μ or ν

[1] We are grateful to Dirk Tasche for pointing this out to us; the first version of the paper contained a more cumbersome constructive proof.

is asymptotically worse. There are however pathological examples of mixing distributions which are not ordered in tail order. For an example we may take $G_\mu(q) = q$ and

$$G_\nu(q) = \begin{cases} 0 & \text{for } q < \frac{1}{2} \\ q - \frac{1}{4}(1-q)^3 \sin(\frac{1}{1-q}) & \text{for } q \geq \frac{1}{2} \end{cases}$$

Obviously the two df's cross infinitely often as $q \to 1$.

Proof of Proposition 6.4. Chose some $q_0 \in (0,1)$ such that $G_\mu(q) \leq G_\nu(q)$ for all $q \in [q_0, 1]$ and moreover $\int_{q_0}^1 (G_\nu(q) - G_\mu(q))dq > 0$. We get by partial integration

$$\int_0^1 q^k dG_\mu(q) = 1 - k \int_0^1 q^{k-1} G_\mu(q) dq$$

and similarly for G_ν. Hence

$$\begin{aligned} \int_0^1 q^k (dG_\mu(q) - dG_\nu(q)) &= k \int_0^{q_0} q^{k-1} \underbrace{(G_\nu(q) - G_\mu(q))}_{\geq -1} dq + k \int_{q_0}^1 q^{k-1} (G_\nu(q) - G_\mu(q)) dq \\ &\geq q_0^{k-1} \left(-q_0 + k \int_{q_0}^1 \left(\frac{q}{q_0}\right)^{k-1} (G_\nu(q) - G_\mu(q)) dq \right) \\ &\geq q_0^{k-1} \left(-q_0 + k \int_{q_0}^1 (G_\nu(q) - G_\mu(q))\, dq \right), \end{aligned}$$

which is obviously positive for k large enough. □

The following Figure 4 which provides moments of various mixing distributions calibrated to have the same first two moments illustrates the previous results.

7 Estimating Default Probabilities and Correlations: Exchangeable Case

In Section 6.2 we have seen that it is useful to be able to calibrate models for homogeneous groups to have known default probabilities and default correlations. This raises the question of how suitable values for the parameters π, π_2 and ρ_Y may be estimated from historical data. In this section we derive a general procedure for estimating π_k in an exchangeable model and make some comments about the asymptotic properties of the estimator. For simplicity we begin with an estimator for π before explaining how the method is generalised to π_k, $k \geq 1$.

7.1 Estimating default probabilities

Assume we can collect historical data on numbers of observed defaults over a specified time period for an exchangeable group of m obligors. If M of these obligors default then the obvious unbiased estimator of π is M/m.

In studying the asymptotics of this estimator as $m \to \infty$ we implicitly assume that we have an infinite exchangeable sequence $Y_1, Y_2, \ldots$ for which an appropriate mixing variable Q with support in $[0,1]$ exists by De Finetti's Theorem. Simple calculations give

$$\begin{aligned} \text{var}(M) &= E(\text{var}(M \mid Q)) + \text{var}(E(M \mid Q)) \\ &= m\left(E(Q) - E(Q^2)\right) + m^2 \text{var}(Q) = m(\pi - \pi_2) + m^2(\pi - \pi^2). \end{aligned}$$

Writing $M^{(m)}$ for the number of defaulting obligors when the portfolio size is m we have that

$$\operatorname{var}\left(\frac{M^{(m)}}{m}\right) \sim \operatorname{var}(Q) = \pi_2 - \pi^2 = \rho_Y \pi(1-\pi), \quad m \to \infty.$$

If $\rho_Y > 0$ (corresponding to $\operatorname{var}(Q) > 0$) then the the variance of the estimator cannot be made arbitrarily small in large portfolios and we do not have a consistent estimator of π as $m \to \infty$. To construct a consistent estimator we require repeated observations of the exchangeable group.

Suppose our time horizon of interest is one year and we have n years of historical data $\{(m_j, M_j),\ j = 1,\dots,n\}$, where m_j denotes the number of obligors observed in year j (treated as deterministic) and M_j is the number of these that default (treated as random). Assume stationarity of the model over time in the sense that there exist identically distributed mixing variables $Q_1,\dots,Q_n$ and defaults in year j are conditionally independent given Q_j. An unbiased estimator of π based on n years of data is

$$\widehat{\pi} := \widehat{\pi}^{(n,m_1,\dots,m_n)} := \frac{1}{n}\sum_{j=1}^{n}\frac{M_j}{m_j}.$$

Consider the behaviour of the estimator $\widehat{\pi}$ as $n \to \infty$. If $Q_1,\dots,Q_n$ are independent mixing variables the variance of the estimator is

$$\operatorname{var}(\widehat{\pi}) = \frac{\pi_2 - \pi^2}{n} + \frac{\pi - \pi_2}{n^2}\sum_{j=1}^{n}\frac{1}{m_j},$$

and the estimator is clearly consistent as $n \to \infty$. For large $m_1,\dots,m_n$ the approximation $\operatorname{var}(\widehat{\pi}) \approx \frac{\pi_2 - \pi^2}{n}$ will be sufficiently accurate and this can be estimated by replacing π and π_2 by suitable estimates (for an estimator of π_2 keep reading). Of course, it might not be realistic to assume independence of the Q_j's due to economic cycles inducing serial dependence.

7.2 Estimating joint default probabilities and default correlation

In a similar fashion to the previous section we seek an unbiased estimator for joint default probabilities π_k which generalises the previous estimator. The following proposition collects the moment results that are required. Consider again a generic exchangeable group of m obligors following an exchangeable Bernoulli mixture model, where M obligors default.

Proposition 7.1. *Define the random variable*

$$\binom{M}{k} := \binom{M}{k}^{(m)} := \begin{cases} \frac{M!}{k!(M-k)!} & 1 \le k \le M, \\ 0 & k > M, \end{cases}$$

to be the number of possible subgroups of k obligors in the M defaulting obligors. Then

$$E\binom{M}{k} = \binom{m}{k}E(Q^k) = \binom{m}{k}\pi_k, \quad 1 \le k \le m, \tag{24}$$

$$\operatorname{var}\binom{M}{k}^{(m)} \sim \binom{m}{k}^2 \operatorname{var}(Q^k) = \binom{m}{k}^2 (\pi_{2k} - \pi_k^2), \quad m \to \infty.$$

Proof. Using the identity

$$\binom{M}{k} = \sum_{i_1,\dots,i_k:\,\{i_1,\dots,i_k\}\subset\{1,\dots,m\}} Y_{i_1}\cdots Y_{i_k},$$

equation (24) is immediate upon taking expectations. Furthermore

$$\begin{aligned}\operatorname{var}\binom{M}{k} &= \operatorname{var}\left(E\left(\binom{M}{k} \mid Q\right)\right)+E\left(\operatorname{var}\left(\binom{M}{k} \mid Q\right)\right)\\ &= \binom{m}{k}^2 \operatorname{var}(Q^k)+E\left(\operatorname{var}\left(\sum Y_{i_1}\cdots Y_{i_k} \mid Q\right)\right), \qquad (25)\end{aligned}$$

and the second term can be calculated to be order $o(m^{2k})$ while the first term is $o(m^{2k+1})$. □

If we have yearly default data we would use these data to build the estimator

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

It is possible (though tedious) to calculate the exact variance of this estimator under an assumption of indendence of $Q_1,\ldots,Q_n$. For example, for $k=2$ we can calculate the exact variance of $\binom{M}{k}/\binom{m}{k}$ in (25) and use this to get

$$\operatorname{var}(\widehat{\pi}_2) = \frac{1}{n^2}\sum_{j=1}^{n}\left(\pi_4\frac{(m_j-2)(m_j-3)}{m_j(m_j-1)}+\pi_3\frac{4(m_j-2)}{m_j(m_j-1)}+\pi_2\frac{2}{m_j(m_j-1)}-\pi_2^2\right).$$

In general, for large $m_1,\ldots,m_n$, the approximation $\operatorname{var}(\widehat{\pi}_k) \approx \frac{\pi_{2k}-\pi_k^2}{n}$ can be used.

Cleary ρ_Y can be estimated by taking $\widehat{\rho}_Y = (\widehat{\pi}_2 - \widehat{\pi}^2)/(\widehat{\pi} - \widehat{\pi}^2)$.

References

Bertsimas, D., and I. Popescu (2000): "Optimal inequalities in probability: a convex optimization approach," preprint, Sloan School, MIT and INSEAD.

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

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

Duffie, D. and Lando, D. (2000): "Term Structure of Credit Risk with Incomplete Accounting Observations," working paper, Graduate School of Business, Stanford University, forthcoming in Econometrica.

Duffie, D. and Singleton, K. (1998): "Simulating Correlated Defaults," working paper, Graduate School of Business, Stanford University.

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.

Feller, W. (1971): *An Introduction to Probability Theory and Its Applications: Volume II.* Wiley, New York.

Frey, R., A. McNeil, and M. Nyfeler (2001): "Credit Risk and Copulas," Preprint, ETH Zürich, available from `http://www.math.ethz.ch/~mcneil`.

Fristedt, B., and L. Gray (1997): *A modern approach to probability theory.* Birkhäuser, Basel.

Goovaerts, M., F. De Vylder, and J. Haezendonck (1984): *Insurance Premiums.* North Holland, Amsterdam.

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

Grandell, J. (1997): *Mixed Poisson Processes.* Chapman and Hall, London.

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

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

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

Li, D. (1999): "On Default Correlation: A Copula Function Approach," working paper, RiskMetrics Group, New York.

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

Mardia, K., J. Kent, and J. Bibby (1979): *Multivariate Analysis.* Academic Press, London.

Marshall, A., and I. Olkin (1960): "A one-sided inequality of the Chebyshew type," *Annals of Mathematical Statistics*, 31, 488–491.

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.

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

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

Wilson, T. (1997): "Portfolio Credit Risk (I & II)," *RISK*, 10(9 & 10).

A Copulas

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

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

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

1. $C(u_1, \dots, u_d)$ is increasing in each component u_i.

2. $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ 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 $a_i \leq b_i$ we have:

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

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 df is by definition a copula, which we denote C. It follows that

$$\begin{aligned} F(x_1, \ldots, x_n) &= P(F_1(X_1) \leq F_1(x_1), \ldots, F_d(X_d) \leq F_d(x_d)) \\ &= C(F_1(x_1), \ldots, F_d(x_d)), \end{aligned} \tag{26}$$

or alternatively $C(u_1, \ldots, u_n) = F(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d))$, where $F_i^{\leftarrow}$ denotes the generalised inverse of the df F_i. Formula (26) 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 continuous margins $F_1, \ldots, F_d$. Then there exists a unique copula $C : [0, 1]^d \to [0, 1]$ such that for all $x_1, \ldots, x_d$ in $\overline{\mathbb{R}} = [-\infty, \infty]$ (26) holds. Conversely, if C is a copula and $F_1, \ldots, F_d$ are distribution functions, then the function F given by (26) is a joint distribution function with margins $F_1, \ldots, F_d$.*

For a proof and extensions to discontinuous marginal distributions we refer to Schweizer and Sklar (1983). Sklar's theorem allows us to define the notion of the copula of a distribution F.

Definition A.3 (Copula of F). If F is a joint df with continuous marginals $F_1, \ldots, F_d$ and (26) holds, we say that C is the 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.

Proposition A.4. *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 $(T_1(X_1), \ldots, T_d(X_d))$ also has copula C.*

Random variables $X_1, \ldots, X_d$ with continuous marginals are independent if and only if their copula is

$$C^{ind}(u_1, \ldots, u_d) = \prod_{i=1}^{d} u_i.$$

Each of $X_1, \ldots, X_d$ is almost surely a strictly increasing function of any of the others (a concept known as comonotonicity) if and only if their copula is

$$C^u(u_1, \ldots, u_d) = \min\{u_1, \ldots, u_d\}.$$

The copula of the d-dimensional Gaussian distribution takes the form

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

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

As in Joe (1997) and Embrechts, McNeil, and Straumann (2001) we define tail dependence as a bivariate concept for pairs of random variables with continuous marginal distributions.

Definition A.5 (Lower Tail Dependence). Let X_1 and X_2 be continuous random variables and $C(u_1, u_2)$ their unique copula. The *coefficient of lower tail dependence* is defined to be $\lambda_\ell = \lim_{u\to 0} \frac{C(u,u)}{u}$. If $\lambda_\ell \in (0,1]$ then the pair (X_1, X_2) (or the copula C) is said to have lower tail dependence. If $\lambda_\ell = 0$ we talk of asymptotic independence in the lower tail.

This definition can be understood by observing that, if X_1 and X_2 have quantile functions $F_1^{\leftarrow}$ and $F_2^{\leftarrow}$ respectively, then

$$\lambda_\ell = \lim_{u\to 0} \frac{C(u,u)}{u} = \lim_{u\to 0} P\left(X_2 \le F_2^{\leftarrow}(u) \mid X_1 \le F_1^{\leftarrow}(u)\right),$$

i.e. the limiting conditional probability X_2 lies below its u-quantile, given that X_1 lies below its u-quantile (or vice versa). For bivariate Gaussian random variables with correlation $\rho < 1$ (and hence the Gaussian copula with parameter ρ) λ_ℓ is zero. Suppose C_ρ^{Ga} denotes the bivariate Gaussian copula with parameter ρ and C^* is some other copula with lower tail dependence. Then, for any $k > 1$ there exists u_0 such that, for $u \le u_0$,

$$\frac{C^*(u,u)}{C_\rho^{\mathrm{Ga}}(u,u)} > k.$$

B Classical Factor Models

Definition B.1 (Classical linear factor model). A d-dimensional random vector $\mathbf{X}$ with mean vector $\boldsymbol{\mu}$ is said to follow a classical p-factor model ($p < d$) if we can write

$$X_i = \sum_{j=1}^{p} a_{i,j}\Theta_j + b_i\varepsilon_i + \mu_i, \quad i = 1, \dots, d, \tag{27}$$

with terms as follows.

1. The $a_{i,j}$ and b_i are real constants.
2. $\boldsymbol{\Theta} = (\Theta_1, \dots, \Theta_p)'$ is a random vector with mean zero (comprising the so-called common factors).
3. $\varepsilon_1, \dots, \varepsilon_d$ are *uncorrelated* random variables with mean zero and variance 1 (comprising the so-called idiosyncratic factors).
4. Θ_j and ε_i are *uncorrelated* for all i, j.

The covariance matrix Σ of $\mathbf{X}$ is clearly given by

$$\Sigma = A\Omega A' + \mathrm{diag}(b_1^2, \dots, b_m^2), \tag{28}$$

where A is an appropriate matrix of constants and Ω is the covariance matrix of $\boldsymbol{\Theta}$. (Some conventions choose the common factors to be uncorrelated so that Ω is the identity matrix.) In fact, it can be shown that a random vector $\mathbf{X}$ has a representation of the form (27) if and only if its covariance matrix Σ can be decomposed as in (28) for some symmetric matrix $\Omega \in \mathbb{R}^{p\times p}$ and some $A \in \mathbb{R}^{d\times p}$ See, for instance, Mardia, Kent, and Bibby (1979).

Note that $\mathbf{X}$ need not be multivariate normally distributed for the factor model to hold. However, if $\mathbf{X}$ is multivariate normal with covariance matrix satisfying (28), then $\boldsymbol{\Theta}, \varepsilon_1, \dots, \varepsilon_d$ can be chosen to be multivariate normal and the word *uncorrelated* can be replaced by *independent* in the definition.

C Pictures

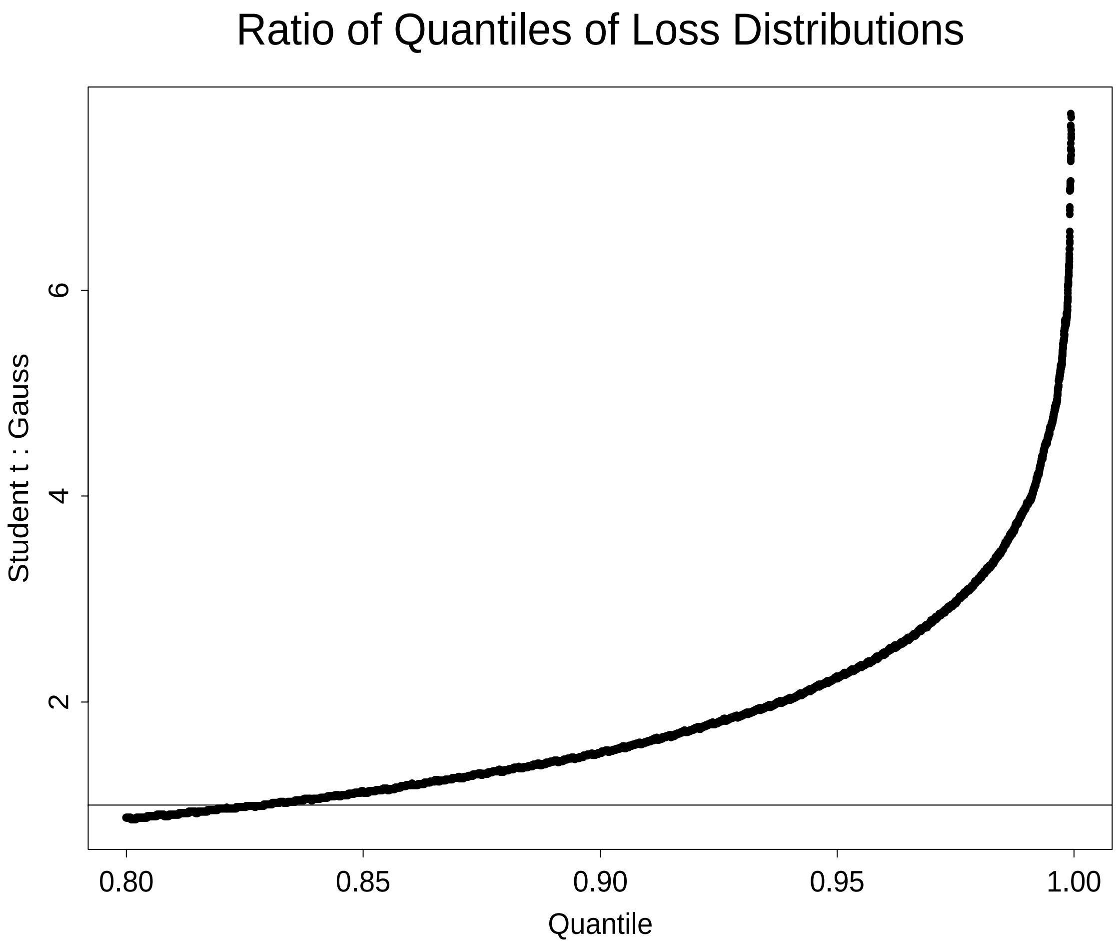

Figure 1: Ratio of estimated quantiles of distribution of M for Student t model with 10 degrees of freedom and Gaussian model in case of Group B with 10000 obligors.

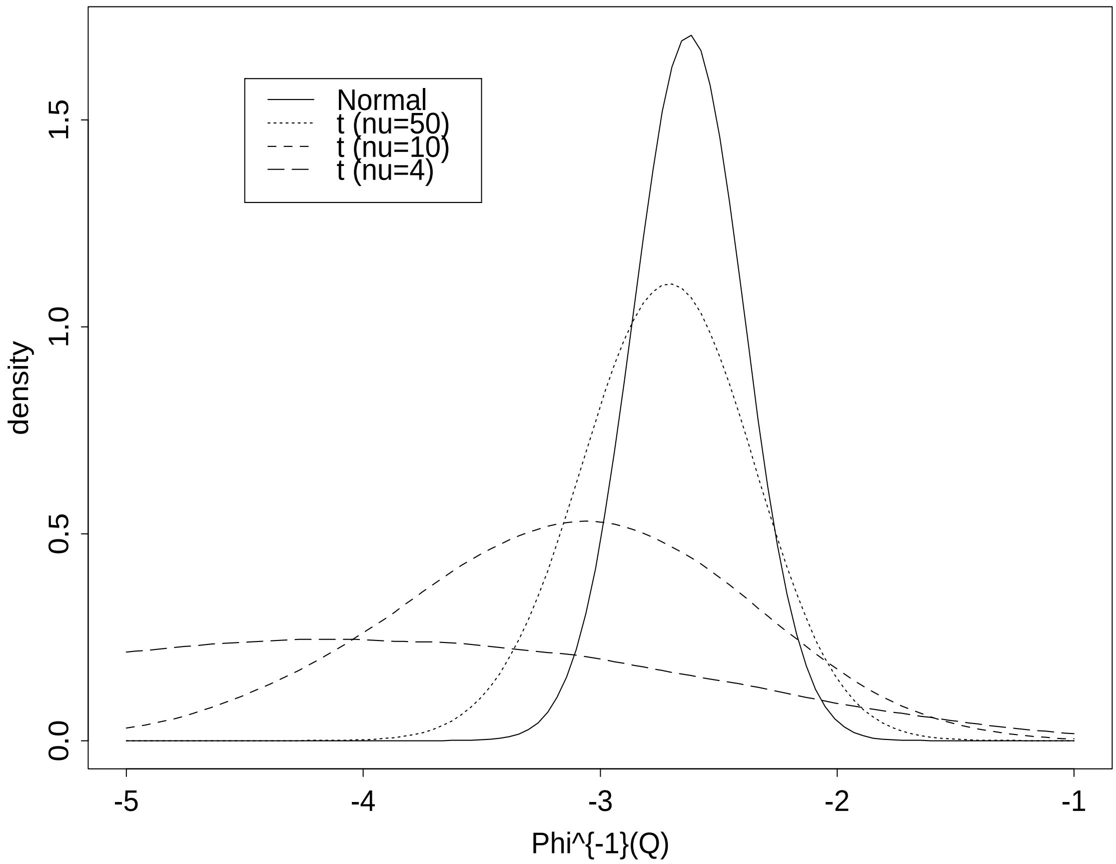

Figure 2: Densities of $\Phi^{-1}(Q)$ for Group B, when latent variables are either Gaussian or Student-t. For all of these distributions $E(Q) = \pi = 0.005$. The heavier the right tail of the distribution of $\Phi^{-1}(Q)$ (or equivalently of Q) the heavier the tail of the distribution of M in sufficiently large portfolios.

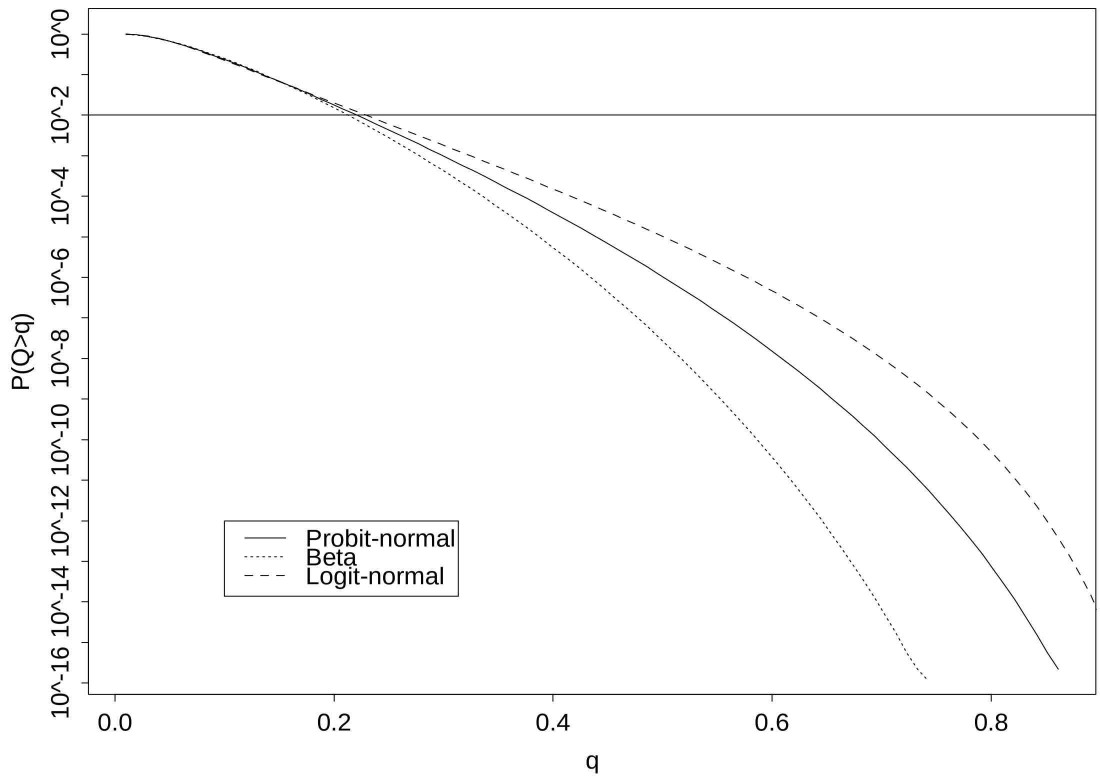

Figure 3: Tail of the mixing distribution G of Q in three different exchangeable Bernoulli mixture models: probit-normal (equivalent to latent variables with Gaussian copula); logit-normal; beta. In all cases the first two moments π and π_2 have the values for group C in Table 2. Horizontal line at 0.01 shows that models only really start to diverge at 99th percentile of mixing distribution.

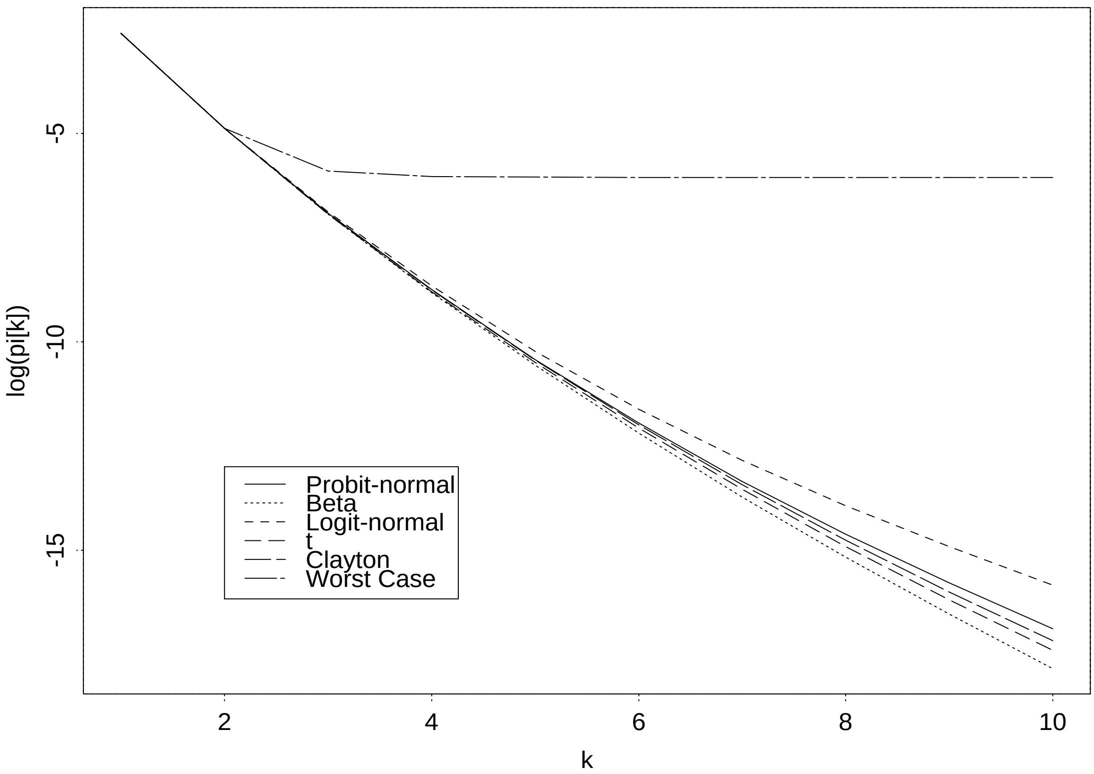

Figure 4: Moments of the mixing distribution G of Q in six different exchangeable Bernoulli mixture models: probit-normal (equivalent to latent variables with Gaussian copula); logit-normal; beta; t (i.e. mixing distribution implied by t copula); Clayton (i.e. mixing distribution implied by Clayton copula); worst-case. In all cases the first two moments π and π_2 have the values for group C in Table 2.
Note that higher moments are ordered in the same way as the survivor functions of the corresponding df's in Figure 3; the worst-case distribution clearly dominates all the other mixing distributions.