MODELING DEPENDENT RISKS WITH MULTIVARIATE ERLANG MIXTURES

BY

SIMON C.K. LEE AND X. SHELDON LIN

ABSTRACT

In this paper, we introduce a class of multivariate Erlang mixtures and present its desirable properties. We show that a multivariate Erlang mixture could be an ideal multivariate parametric model for insurance modeling, especially when modeling dependence is a concern.When multivariate losses are governed by a multivariate Erlang mixture, many quantities of interest such as joint density and Laplace transform, moments, and Kendall's tau have a closed form. Further, the class is closed under convolutions and mixtures, which enables us to model aggregate losses in a straightforward way. We also introduce a new concept called quasi-comonotonicity that can be useful to derive an upper bound for individual losses in a multivariate stochastic order and upper bounds for stop-loss premiums of the aggregate loss. Finally, an EM algorithm tailored to multivariate Erlang mixtures is presented and numerical experiments are performed to test the efficiency of the algorithm.

KEYWORDS

Erlang mixture, dependent risk, multivariate analysis, quasi-comonotonicity, aggregate losses, EM algorithm.

1. INTRODUCTION

Modeling dependent financial and insurance risks is central to the sound risk management of an insurance company. Multivariate parametric distributions were often used for this purpose in the past. In recent years, the use of copulas has become a dominant choice for multivariate modeling in finance, insurance, and even statistics. See, for example, Frees and Valdez (1998), Genest et al. (2009) and the papers in a special issue (Volume 44, Issue 2) of *Insurance Mathematics and Economics* and references therein. An advantage of the copula approach is that it uses a two stage procedure that separates the dependence structure of a model distribution from its marginals. Many researchers find this feature very appealing as they make use of the rich sources

Astin Bulletin 42(1), 153-180. doi: 10.2143/AST.42.1.2160739

of univariate modeling techniques. However, the use of copulas has some shortcomings. As Harry Joe pointed out in Joe (1997), p. 84, an ideal multivariate parametric model should have the following desirable properties:

A. interpretability, which could mean something like mixture, stochastic or latent variable representation;
B. the closure property under the taking of margins, in particular the bivariate margins belonging to the same parametric family (this is especially important if, in statistical modeling, one thinks first about appropriate univariate margins, then bivariate and sequentially to higher-order margins);
C. a flexible and wide range of dependence (with type of dependence structure depending on applications);
D. a closed-form representation of the cdf and density (a closed-from cdf is useful if the data are discrete and a continuous random vector is used), and if not closed-form, then a cdf and density that are computationally feasible to work with.

For the commonly used copulas, Properties C and D are some time not satisfied due to the specific analytical form of those copulas. Dimensionality could be another potential problem. Although this is not unique to copulas, it seems that some copulas could make the problem worse. See Mikosch (2006) for some criticisms on the copula methodology.

In this paper, we propose a direct approach by using a class of multivariate distributions that we will call multivariate Erlang mixtures. A distribution in the class is a mixture such that each of its component distributions is the joint distribution of independent Erlangs (gamma with integer shape parameter) which share a common scale parameter. Property A is partly addressed as the mixing distribution of the mixture gives good indication about tail heaviness and mode positions. We will show that Properties B, C and D are satisfied: all the marginals and conditional marginals remain to be in the same class; the class is dense in the sense of weak convergence in the space of all positive continuous multivariate distributions; and joint cdf and density, moments and Laplace transform, as well as correlation measures, are of a closed-form. Furthermore, a simple and stable EM algorithm is available to fit a multivariate Erlang mixture to multivariate positive data. The algorithm can easily handle data sets with very high dimension.

The paper is organized as follows. We introduce the class of multivariate Erlang mixtures and present its basic properties in Section 2. We also show that for any positive continuous multivariate distribution, there is a sequence of multivariate Erlang mixtures that converge to the distribution weakly. We then derive the moments and common measures of association in Section 3. In Section 4, we introduce a new concept called quasi-comonotonicity that is useful to derive an upper bound of individual losses in a multivariate stochastic order. Section 5 covers aggregate and excess losses, in which we derive the distribution of the aggregate loss and analytical formulas for associated risk measures such as value at risk (VaR) and tail VaR. We then give sharp upper bounds for the variance and stop-loss premium of the aggregate loss and derive the distribution of the multivariate excess losses. We present an expectation-maximization (EM) algorithm tailored to the class of multivariate Erlang mixtures in Section 6. To show the efficiency of the EM algorithm, we fit the data generated from a multivariate lognormal distribution of 12 dimensions and test its goodness of fit in Subsection 6.2 and present two bivariate Erlang mixtures with extreme dependence in Subsection 6.3. We conclude by making some remarks on potential applications of the multivariate Erlang mixture in Section 7.

2. Multivariate Erlang Mixtures

In this section, we introduce the class of multivariate Erlang mixtures with common scale parameter and study its properties.

Let $p(x; m, \theta)$ be the density of the Erlang distribution with shape parameter m, a positive integer, and scale parameter θ. That is,

$$p(x; m, \theta) = \frac{x^{m-1} e^{-x/\theta}}{\theta^m (m-1)!}, \quad x > 0. \tag{2.1}$$

Define the density of a k-variate Erlang mixture as

$$f(\mathbf{x} \,|\, \theta, \boldsymbol{\alpha}) = \sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \alpha_{\mathbf{m}} \prod_{j=1}^{k} p(x_j; m_j, \theta). \tag{2.2}$$

In (2.2) and hereafter, we always denote $\mathbf{x} = (x_1, \ldots, x_k)$, $\mathbf{m} = (m_1, \ldots, m_k)$, $\boldsymbol{\alpha} = (\alpha_{\mathbf{m}}; m_i = 1, 2, \ldots; i = 1, 2, \ldots, k)$ with each $\alpha_{\mathbf{m}} \geq 0$ and $\sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \alpha_{\mathbf{m}} = 1$.

The proposed class of multivariate Erlang mixtures has many desirable properties. One of them is the conditional independence structure of the distributions, which allows us to easily calculate many quantities of interest in insurance. It is also easy to verify that this class is closed under mixtures. However, before considering the use of a multivariate Erlang mixture for insurance modeling and valuation purposes, we must ask ourselves whether such a model can fit insurance loss data of any kind well, a stronger assumption than Property C stated in Section 1. The following theorem provides a theoretical justification for such a possibility. More precisely, it shows that for any positive continuous multivariate distribution, we may construct a sequence of multivariate Erlang mixtures that converge to the target distribution weakly.

Theorem 2.1. *The class of multivariate Erlang mixtures of form (2.2) is dense in the space of positive continuous multivariate distributions in the sense of weak convergence.*

The proof is given in Appendix A.

It is well known that a distribution is a univariate Erlang mixture if and only if it can be expressed as a compound exponential distribution (see Lee and Lin (2010) and references therein). We show in the following theorem that this property can be extended to the class of multivariate Erlang mixtures. As a result, various properties on the marginal distributions can be derived subsequently.

Theorem 2.2. *If* $\mathbf{X} = (X_1, \ldots, X_k)$ *has a multivariate Erlang mixture of (2.2), then each marginal random variable* X_j *has a compound exponential distribution, i.e.* $X_j = \sum_{i=1}^{N_j} E_{ij}$ *in distribution for* $j = 1, \ldots, k$*, where* N_j *is the primary counting random variable and* E_{ij}*,* $i = 1, 2, \ldots, j = 1, \ldots, k$*, are iid exponential random variables with mean* θ*. Moreover, the joint primary distribution of* $\mathbf{N} = (N_1, \ldots, N_k)$ *has probability function*

$$P(\mathbf{N} = \mathbf{m}) = \alpha_{\mathbf{m}}, m_j = 1, 2, \ldots; \; j = 1, \ldots, k. \tag{2.3}$$

Proof. Suppose that random variables $X_1, \ldots, X_k$ are of the from: $X_j = \sum_{i=1}^{N_j} E_{ij}$, $j = 1, \ldots, k$. Then, the characteristic function of $\mathbf{X}$ is given by

$$\begin{aligned}
\varphi(\mathbf{z}) &= E\{e^{\mathbf{i}\mathbf{z}\cdot\mathbf{X}}\} = E\left\{E\{e^{\mathbf{i}\mathbf{z}\cdot\mathbf{X}} \mid \mathbf{N}\}\right\} \\
&= E\left\{E\left\{\prod_{j=1}^{k} e^{\mathbf{i}z_j \sum_{i=1}^{N_j} E_{ij}} \middle| \mathbf{N}\right\}\right\} \\
&= E\left\{\prod_{j=1}^{k} E\left\{e^{\mathbf{i}z_j \sum_{i=1}^{N_j} E_{ij}} \middle| \mathbf{N}\right\}\right\} \\
&= E\left\{\prod_{j=1}^{k} \left(\frac{1}{1 - \mathbf{i}\theta z_j}\right)^{N_j}\right\} \\
&= P_{\mathbf{N}}\left(\frac{1}{1 - \mathbf{i}\theta z_1}, \ldots, \frac{1}{1 - \mathbf{i}\theta z_k}\right).
\end{aligned} \tag{2.4}$$

where $P_{\mathbf{N}}(\mathbf{z})$ is the probability generating function of $\mathbf{N}$. It is obvious that the last expression is the characteristic function of the multivariate Erlang mixture (2.2). The uniqueness of the characteristic function leads to the theorem. □

Corollary 2.1. *The marginal distribution of* X_j *is a univariate Erlang mixture. Moreover, the weights of the mixture are*

$$\alpha_{m_j}^{(j)} \overset{\text{def}}{=} \sum_{m_l, l \neq j; m_l = 1, 2, \ldots} \alpha_{\mathbf{m}}, \quad m_j = 1, 2, \ldots.$$

Furthermore, any p-variate $(p < k)$ *marginal is a p-variate Erlang mixture.*

Proof. It is obvious. □

https://doi.org/10.2143/AST.42.1.2160739 Published online by Cambridge University Press

Corollary 2.2. *(Covariance Invariance) The covariance of any marginal pair* (X_j, X_l) *is proportional to the covariance of* (N_j, N_l). *More precisely,*

$$Cov(X_j, X_l) = \theta^2 Cov(N_j, N_l). \tag{2.6}$$

Proof. It is obvious that $E(X_j) = \theta E(N_j)$, $j = 1, \ldots, k$. Since for $j \neq l$,

$$\begin{aligned} E(X_j X_l) &= E\left\{\left[\sum_{i=1}^{N_j} E_{ij}\right]\left[\sum_{i=1}^{N_l} E_{il}\right]\right\} \\ &= E\left\{E\left\{\left[\sum_{i=1}^{N_j} E_{ij}\right]\left[\sum_{i=1}^{N_l} E_{il}\right]\middle| N_j, N_l\right\}\right\} \\ &= E\left\{\left[\theta N_j\right]\left[\theta N_l\right]\right\} = \theta^2 E(N_j N_l), \end{aligned}$$

we have

$$\begin{aligned} Cov(X_j, X_l) &= E(X_j X_l) - E(X_j) E(X_l) = \theta^2 E(N_j N_l) - \theta^2 E(N_j) E(N_l) \\ &= \theta^2 Cov(N_j, N_l). \end{aligned}$$

□

Corollary 2.3. *The marginal random variables* $X_1, \ldots, X_k$ *are mutually independent if the counting random variables* $N_1, \ldots, N_k$ *are mutually independent. In this case, we have*

$$\alpha_{\mathbf{m}} = \prod_{j=1}^{k} \alpha_{m_j}^{(j)}$$

Proof. It is obvious. □

Univariate phase-type distributions have been widely used in modeling insurance losses in insurance risk theory. See Asmussen and Albrecher (2010) and references therein. The main advantage of using a univariate phase-type distribution is the applicability of the matrix-analytic method. As a result, many quantities of interest related to the time of ultimate ruin of an insurance risk models can be written in an analytical form. Assaf et al. (1984) introduced a class of multivariate phase-type distributions. It is a generalization of the class of univariate phase-type distributions. They also showed that the class has similar desirable properties, as in the univariate case: the matrix-analytic method is applicable, the density, moments and Laplace transform of a multivariate phase-type distribution can be written in an analytical form, and the class is dense in the space of all multivariate positive distributions. In the following theorem, we show that the class of finite multivariate Erlang mixtures is a subclass.

Theorem 2.3. *A finite multivariate Erlang mixture is a multivariate phase-type distribution defined in Assaf et al. (1984).*

Proof. Obviously, an Erlang distribution is a univariate phase-type distribution. Thus, it follows from Theorem 2.1 of Assaf et al. (1984) that $\prod_{j=1}^{k} p(x_j, m_j, \theta)$ is a multivariate phase-type distribution. Since the class of multivariate phase-type distributions is closed under finite mixtures (Theorem 2.2 of Assaf et al. (1984), a finite multivariate Erlang mixture is a multivariate phase-type distribution. □

Although a phase-type distribution has many desirable analytical properties, fitting it to data is challenging. Even in the univariate case, the non-uniqueness of phase-type representation and exponential increase in the number of parameters make any estimation procedure inefficient (Asmussen et al. (1996)). To the best of our knowledge, there is no estimation procedure available for multivariate phase-type distributions higher than 2 dimensions. The bivariate phase-type modeling can be found in Eisele (2005). Readers can easily find that the model cannot allow for a large number of states which are often required when modeling heavy tailed data. Furthermore, it is extremely difficult to construct a phase-type representation with given marginals and predetermined dependent structure except the independent case. This might be the reason why few applications of multivariate phase-type distributions can be found. On the other hand, the class of multivariate Erlang mixtures is very easy to manipulate and analyze. It can be easily used to fit loss data using an EM algorithm as we will show in Section 6.1. Moreover, the conditional independence structure allows us to deal with high dimensional data. In other words, the curse of dimensionality in terms of computational complexity can be partly overcome.

3. Moments and Common Measures of Association

In this section, we show that, as in the univariate case, the moments of a multivariate Erlang mixture can be easily written in a closed form. Furthermore, some commonly used measures of association such as Kendall's tau and Spearman's rho can also be written explicitly.

Theorem 3.1. *Let* $\mathbf{X} = (X_1, \ldots, X_k)$ *have the multivariate Erlang mixture (2.2). Then, the joint moment*

$$E\left\{\prod_{j=1}^{k} X_j^{n_j}\right\} = \theta^n \sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \alpha_{\mathbf{m}} \prod_{j=1}^{k} \frac{(m_j + n_j - 1)!}{(m_j - 1)!}, \tag{3.1}$$

where $n = \sum_{j=1}^{k} n_j$.

Proof. It simply follows from that the n_j-th moment of $p(x; m_j, \theta)$ is $\frac{(m_j+n_j-1)!}{(m_j-1)!}\theta^{n_j}$ and the conditional independence. □

https://doi.org/10.2143/AST.42.1.2160739 Published online by Cambridge University Press

In the following, we derive explicit formulas for Kendall's tau and Spearman's rho, two most widely used measures of association, for the multivariate Erlang mixture (2.2). Without the loss of generality, we assume $k = 2$. That is, we consider a bivariate Erlang mixture with mixing weights α_{ij}, $i, j = 1, 2, \ldots$

The population version of Kendall's tau for a pair of continuous random variables X and Y measures the tendency that X and Y will move in the same direction (concordance) and it is defined as

$$\tau = P\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_2) < 0\}, \qquad (3.2)$$

where (X_1, Y_1) and (X_2, Y_2) are two iid copies of (X, Y). Unlike the (Pearson's) correlation coefficient, it does not assume linear relationship. In this regard, Kendall's tau is more meaningful in measuring the correlation between two random variables. The population version of Spearman's rank correlation coefficient (Spearman's rho) is another commonly used measure of association. It is defined as

$$\rho = 3(P\{(X_1 - X_2)(Y_1 - Y_3) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_3) < 0\}), \qquad (3.3)$$

where (X_1, Y_1), (X_2, Y_2) and (X_3, Y_3) are iid copies of (X, Y). It is known that the value of the both ranges from –1 to 1: 1 indicates a perfect agreement and –1 a perfect disagreement.

Theorem 3.2. *Kendall's tau of a bivariate Erlang mixture is given by*

$$\tau = 4 \sum_{i,j=0}^{\infty} \sum_{k,l=1}^{\infty} \binom{i+k-1}{i} \binom{j+l-1}{j} \frac{Q_{ij}\,\alpha_{kl}}{2^{i+j+k+l}} - 1, \qquad (3.4)$$

where $Q_{ij} = \sum_{k=i+1}^{\infty} \sum_{l=j+1}^{\infty} \alpha_{kl}$ *is the survival function of the mixing distribution.*

Proof. Let $\bar{F}(x, y)$ be the survival function for (X, Y) and $\bar{P}(x; i, \theta)$ the survival function of $p(x; i, \theta)$. Then

$$\begin{aligned} \bar{F}(x, y) &= \sum_{k,l=1}^{\infty} \alpha_{kl}\, \bar{P}(x; k, \theta)\, \bar{P}(y; l, \theta) \\ &= \sum_{k,l=1}^{\infty} \alpha_{kl}\, e^{-\frac{x+y}{\theta}} \sum_{i,j=0}^{k-1,\, l-1} \frac{x^i y^j}{\theta^{i+j} i! j!} \\ &= e^{-\frac{x+y}{\theta}} \sum_{i,j=0}^{\infty} Q_{ij} \frac{x^i y^j}{\theta^{i+j} i! j!}, \end{aligned} \qquad (3.5)$$

where $\sum_{i,j=0}^{k,l} = \sum_{i=0}^{k} \sum_{j=0}^{l}$. It is well known that

https://doi.org/10.2143/AST.42.1.2160739 Published online by Cambridge University Press

$$\tau = 4\int_0^\infty\int_0^\infty \bar{F}(x,y)f(x,y)\,dxdy - 1.$$

Thus,

$$\begin{aligned}\tau &= 4\int_0^\infty\int_0^\infty e^{-\frac{x+y}{\theta}} \sum_{i,j=0}^\infty Q_{ij}\frac{x^i y^j}{\theta^{i+j} i!j!} \sum_{k,l=1}^\infty \alpha_{kl}\, e^{-\frac{x+y}{\theta}} \frac{x^{k-1}y^{l-1}}{\theta^{k+l}(k-1)!(l-1)!}\,dxdy - 1 \\ &= 4\sum_{i,j=0}^\infty \sum_{k,l=1}^\infty Q_{ij}\,\alpha_{kl}\int_0^\infty\int_0^\infty e^{-\frac{2(x+y)}{\theta}} \frac{x^{i+k-1}y^{j+l-1}}{\theta^{i+j+k+l} i!j!(k-1)!(l-1)!}\,dxdy - 1 \\ &= 4\sum_{i,j=0}^\infty \sum_{k,l=1}^\infty \binom{i+k-1}{i}\binom{j+l-1}{j}\frac{Q_{ij}\,\alpha_{kl}}{2^{i+j+k+l}} - 1. \end{aligned} \tag{3.6}$$

□

Theorem 3.3. *Spearman's rho of a bivariate Erlang mixture is given by*

$$\rho = 12\sum_{i,j=0}^\infty \sum_{k,l=1}^\infty \binom{i+k-1}{i}\binom{j+l-1}{j}\frac{Q_{ij}\,\alpha_k^{(1)}\alpha_l^{(2)}}{2^{i+j+k+l}} - 3 \tag{3.7}$$

Proof. The derivation is similar and omitted. □

4. Quasi-Comonotonicity

Comonotonicity is an important concept in studying the dependence structure of multivariate risks with the given/fixed marginal distribution of the risks. It has applications in worst scenario analysis in insurance and finance. If losses from an insurance portfolio are comonotonic, they exhibit the strongest positive dependence among themselves and thus are undiversifiable. This type of insurance portfolios is the least favorable to an insurer. For the investigation of comonotonic risk and its applications in actuarial science and finance, see Dhaene et al. (2002a), Dhaene et al. (2002b) and references therein.

The concept of comonotonicity might not be best suitable for the class of multivariate Erlang mixtures. Given that the marginals are univariate Erlang mixtures, the comonotonic multivariate distribution is in general not a multivariate Erlang mixture. That motivates us to introduce a very similar concept called quasi-comonotonicity in this section such that a quasi-comonotonic multivariate distribution belongs to the class of multivariate Erlang mixtures. As we will show in this and next sections, quasi-comonotonicity is essentially an equivalent to comonotonicity except that it is defined within the class and almost all the properties associated with comonotonicity hold.

We begin by introducing a multivariate stochastic order called the upper orthant order. See Shaked and Shanthikumar (1994), p. 140, or Joe (1997), p. 36.

Definition 4.1 (Upper orthant order). *Let* $\mathbf{X} = (X_1, \ldots, X_k)$ *and* $\mathbf{Y} = (Y_1, \ldots, Y_k)$ *be two random vectors with survival functions* $\bar{F}(x_1, \ldots, x_k)$ *and* $\bar{G}(x_1, \ldots, x_k)$. *We say that* $\mathbf{X}$ *is smaller than* $\mathbf{Y}$ *in the upper orthant order (denoted by* $\mathbf{X} \leq_{uo} \mathbf{Y}$*) if, for any* $\mathbf{x} = (x_1, \ldots, x_k)$,

$$\bar{F}(\mathbf{x}) \leq \bar{G}(\mathbf{x}). \tag{4.1}$$

The upper orthant order is a generalization of the usual univariate stochastic order. Related results can be found in Shaked and Shanthikumar (1994) and Denuit et al. (2005). In the following theorem, we show that the upper orthant order between two mixing distributions is transformable to the upper orthant order between the two corresponding multivariate Erlang mixtures.

Theorem 4.1. *Suppose that random vectors* $\mathbf{X} = (X_1, \ldots, X_k)$ and $\mathbf{Y} = (Y_1, \ldots, Y_k)$ *have a multivariate Erlang mixture, and* $\mathbf{N} = (N_1, \ldots, N_k)$ and $\mathbf{M} = (M_1, \ldots, M_k)$ *are the corresponding counting random vectors, respectively. If* $\mathbf{N} \leq_{uo} \mathbf{M}$, *then* $\mathbf{X} \leq_{uo} \mathbf{Y}$.

Proof. Denote the survival functions of $\mathbf{X}, \mathbf{Y}, \mathbf{N}$ and $\mathbf{M}$ by $\bar{F}(\mathbf{x})$, $\bar{G}(\mathbf{x})$, $\bar{H}(\mathbf{m})$ and $\bar{I}(\mathbf{m})$, respectively. Then, we have

$$\bar{H}(\mathbf{m}) \leq \bar{I}(\mathbf{m}),$$

for all positive integer vectors $\mathbf{m}$. Thus,

$$\begin{aligned}
\bar{F}(\mathbf{x}) &= \sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \alpha_{\mathbf{m}}^{H} \prod_{j=1}^{k} \sum_{i=0}^{m_j-1} \frac{x_j^i e^{-x_j/\theta}}{\theta^i i!} \\
&= \sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \bar{H}(\mathbf{m}) \prod_{j=1}^{k} \frac{x_j^{m_j-1} e^{-x_j/\theta}}{\theta^{m_j-1}(m_j-1)!} \\
&\leq \sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \bar{I}(\mathbf{m}) \prod_{j=1}^{k} \frac{x_j^{m_j-1} e^{-x_j/\theta}}{\theta^{m_j-1}(m_j-1)!} = \bar{G}(\mathbf{x}).
\end{aligned} \tag{4.2.}$$

By the definition of the upper orthant order, $\mathbf{X} \leq_{uo} \mathbf{Y}$. □

We now focus on multivariate Erlang mixtures with fixed marginals. Borrowing the notation from Joe (1997), let $\mathcal{F}_\theta(F_1, \ldots, F_k)$ be the subclass of multivariate Erlang mixtures with common scale parameter θ in which each of the distributions has marginal cdf's $F_1(x_1), \ldots, F_k(x_k)$.

The Fréchet-Hoeffding upper and lower bounds (Hoeffding (1940), Hoeffding (1941), Fréchet (1951) and Fréchet (1958)) are defined as

$$\begin{aligned}
F_U(\mathbf{x}) &= \min\{F_1(x_1), \cdots, F_k(x_k)\}, \\
F_L(\mathbf{x}) &= \max\{F_1(x_1) + \cdots + F_k(x_k) - k + 1, 0\}.
\end{aligned} \tag{4.3}$$

The Fréchet-Hoeffding upper bound is a cdf but the Fréchet-Hoeffding lower bound is not always a cdf except when $k = 2$ (Joe (1997), pp. 57-8). It follows from Theorem 3.5 of Joe (1997) that the Fréchet-Hoeffding upper bound is an upper bound of the subclass $\mathcal{F}_\theta(F_1, \ldots, F_k)$ in the upper orthant order and the Fréchet-Hoeffding lower bound, if it is a cdf, is a lower bound of the subclass in the upper orthant order. Furthermore, the random vector with $F_U(\mathbf{x})$ is comonotonic and vice versa. Similarly, the random vector with $F_L(\mathbf{x})$ is counter-comonotonic, when it is a cdf. See Dhaene et al. (2002a).

Although many useful risk measures for the Fréchet-Hoeffding bounds in (4.3) can be computed easily (see Dhaene et al. (2006)), the analytical form of these multivariate distributions might be difficult to obtain. More importantly, the Fréchet-Hoeffding bounds are in general not a multivariate Erlang mixture. Thus the Fréchet-Hoeffding bounds in the upper orthant order are not sharp bounds. As a result, the bounds based on the Fréchet-Hoeffding bounds for the variance and stop-loss premiums of the corresponding aggregate loss are not sharp either. To avoid these shortcomings and maintain most of the desirable properties of the Fréchet-Hoeffding bounds at meantime, we introduce the concept of quasi-comonotonicity.

Definition 4.2. *Let random vector* $\mathbf{X} = (X_1, \ldots, X_k)$ *have a multivariate Erlang mixture with marginal cdf's* $F_1(x_1), \ldots, F_k(x_k)$, *and* $\mathbf{N} = (N_1, \ldots, N_k)$ *be the corresponding counting random vector. If* $\mathbf{N}$ *is comonotonic, we say that* $\mathbf{X}$ *is quasi-comonotonic. In this case, we denote the comonotonic counting random vector by* $\mathbf{N^U} = (N_1^U, \ldots, N_k^U)$*, the quasi-comonotonic random vector by* $\mathbf{X^U} = (X_1^U, \ldots, X_k^U)$ *and its cdf by* $F_{QU}(\mathbf{x})$*. The notation is slightly different from that in Dhaene et al. (2002a).*

Similarly, if $\mathbf{N}$ *is countercomonotonic, we say that* $\mathbf{X}$ *is quasi-countercomonotonic. Similar notation applies:* $\mathbf{N^L} = (N_1^L, \ldots, N_k^L)$, $\mathbf{X^L} = (X_1^L, \ldots, X_k^L)$ and $F_{QL}(\mathbf{x})$.

Obtaining the quasi-comonotonic distribution and the quasi-countercomonotonic distribution is fairly straightforward. Suppose that we are given k univariate Erlang mixtures $F_1(x_1), \ldots, F_k(x_k)$ as the marginals. The cdf of the mixing distribution of each $F_j(x_j)$, $j = 1, \ldots, k$, can easily be obtained. We then may apply the formulas in (4.3) to obtain the multivariate mixing distributions for the quasi-comonotonic and quasi-countercomonotonic multivariate Erlang mixtures, respectively. In fact, we do not even need to use the formulas. There is a simple recursive procedure for counting distributions as demonstrated in an example in the end of this section. We would like to remark that similar to the Fréchet-Hoeffding lower bound $F_L(\mathbf{x})$, when $k \geq 3$ there is no guarantee that $F_{QL}(\mathbf{x})$ is a cdf.

In the following we will show that the quasi-comonotonic distribution and the quasi-countercomonotonic distribution may serve as the sharp upper and lower bounds.

Theorem 4.2. *Let* $F_{QU}(\mathbf{x})$ *be the quasi-comonotonic cdf, and* $F_{QL}(\mathbf{x})$ *be the quasi-countercomonotonic cdf (if it exists) with fixed marginals* $F_1(x_1), \ldots, F_k(x_k)$,

https://doi.org/10.2143/AST.42.1.2160739 Published online by Cambridge University Press

respectively. Then, they are the sharp upper and lower bounds for all the distributions in $\mathcal{F}_\theta(F_1, \ldots, F_k)$ *in the upper orthant order. In other words, for any* $F \in \mathcal{F}_\theta(F_1, \ldots, F_k)$,

$$\bar{F}_{QL}(\mathbf{x}) \leq \bar{F}(\mathbf{x}) \leq \bar{F}_{QU}(\mathbf{x}),\ \mathbf{x} > 0. \tag{4.4}$$

Proof. The proof is straightforward. It follows from Theorem 3.5 of Joe (1997) that for any $\mathbf{N}$ that generates a multivariate Erlang mixture in $\mathcal{F}_\theta(F_1, \ldots, F_k)$, we have

$$\mathbf{N^L} \leq_{uo} \mathbf{N} \leq_{uo} \mathbf{N^U}.$$

Inequality (4.4) follows immediately from Theorem 4.1. The sharpness is obvious as the bounds are the members of the subclass. □

The quasi-comonotonic random vector and the quasi-countercomonotonic random vector are not only upper and lower bounds in the upper orthant order but also upper and lower bounds for the pairwise correlation coefficients as seen in the following corollary.

Corollary 4.1. *For any* $\mathbf{X}$ *with its cdf in* $\mathcal{F}_\theta(F_1, \ldots, F_k)$*, we have*

$$Corr(X_j^L, X_l^L) \leq Corr(X_j, X_l) \leq Corr(X_j^U, X_l^U),\ j \neq l. \tag{4.5}$$

Proof. Without the loss of generality, assume $j = 1$, $l = 2$. (4.5) follows from (4.4) and

$$E\{X_1 X_2\} = \int_0^\infty \int_0^\infty \bar{F}(x_1, x_2, 0, \cdots, 0)\, dx_1\, dx_2.$$

□

We now demonstrate how to obtain the quasi-comonotonic distribution and the quasi-countercomonotonic distribution using a simple example. As shown in this example, finding the distributions is simple. The recursive method we use in the example can be extended to any dimension.

Example 4.1 (A Simple Example). In this example, we are given three marginal densities:

$$\begin{aligned} f(x) &= 0.4\, p(x;\, 10,\, 100) + 0.4\, p(x;\, 30,\, 100) + 0.2\, p(x;\, 80,\, 100) \\ g(y) &= 0.4\, p(y;\, 20,\, 100) + 0.3\, p(y;\, 40,\, 100) + 0.3\, p(y;\, 70,\, 100) \\ h(z) &= 0.2\, p(z;\, 4,\, 100) + 0.5\, p(z;\, 5,\, 100) + 0.3\, p(z;\, 6,\, 100) \end{aligned} \tag{4.6}$$

Note that the shape parameters are arranged in the ascending order.

We first derive the density of the upper bound $f_{QU}(x, y, z)$: all the possible combinations and their weights. They can be determined in a tabulation method as shown in Table 4.1: Start with the Erlangs with the smallest shape parameter in each of the marginals. In the first row, list these Erlangs $(10, 20, 4)$. The respective weights are 0.4, 0.4 and 0.2. Assign the smallest weight (0.2) to this combination. Thus, the 'unused' weights of Erlangs with shape parameters 10 and 20 are 0.2 and 0.2 respectively. Since the weight for the Erlang with shape parameter 4 is 'used', the Erlang is removed. The Erlangs with the smallest shape parameter in each of the marginals are now 10, 20, 5 with weights 0.2, 0.2 and 0.5, respectively. They are listed in the second row and the smallest weight (0.2) is assigned to the combination. Since the weights of Erlangs with shape parameters 10 and 20 are 'used', these two Erlangs are removed. The Erlangs with the smallest shape parameter in each of the marginals are (30, 40, 5) with weights 0.4, 0.3 and 0.3, respectively. Thus, 0.3 is assigned to this combination and they are list in the next row. Continue in this matter until we exhaust all the Erlangs.

TABLE 4.1.

WEIGHTS AND COMBINATIONS FOR $f_{QU}(x, y, z)$

Weights	x	y	z
0.2	10	20	4
0.2	10	20	5
0.3	30	40	5
0.1	30	70	6
0.2	80	70	6

Based on Table 4.1, we have

$$\begin{aligned} f_{QU}(x,y,z) = {} & 0.2\, p(x; 10, 100)\, p(y; 20, 100)\, p(z; 4, 100) + 0.2\, p(x; 10, 100) \\ & p(y; 20, 100)\, p(z; 5, 100) \\ & + 0.3\, p(x; 30, 100)\, p(y; 40, 100)\, p(z; 5, 100) + 0.1\, p(x; 30, 100) \\ & p(y; 70, 100)\, p(z; 6, 100) \\ & + 0.2\, p(x; 80, 100)\, p(y; 70, 100)\, p(z; 6, 100). \end{aligned}$$

The lower bound is computed by using (4.3) on the mixing distributions. Since we are essentially dealing with counting distributions, the Fréchet-Hoeffding lower bound can be obtained easily. The weights are then the joint probability function that can be computed using a formula in Example 2.21 of Nelson (1999). The weights and combinations are given in Table 4.2. The density of the trivariate Erlang mixture of 13 components can be written accordingly. It is easy to check that the joint distribution reproduces the three marginals. However, one can see that this is not a proper density as some weights are negative. As pointed out in Joe (1997), when there are more than 2 marginals, the Fréchet-Hoeffding lower bound is often not a proper distribution, which is also the case here.

TABLE 4.2.

WEIGHTS AND COMBINATIONS FOR $f_{QL}(x,y,z)$

Weights	x	y	z
0.1	10	40	6
0.1	10	70	5
0.2	10	70	6
0.2	30	20	6
0.2	30	40	5
0.2	30	70	5
–0.2	30	70	6
0.1	80	20	5
0.1	80	20	6
0.1	80	40	5
–0.1	80	40	6
0.2	80	70	4
–0.2	80	70	5

The upper and lower bounds $F_{QU}(\mathbf{x})$ and $F_{QL}(\mathbf{x})$ may be used to model dependency for a portfolio of k correlated losses. Suppose that these losses have cdf's $F_1(x_1), \ldots, F_k(x_k)$, and they are positive correlated at a level of $0 \le \gamma \le 1$. One way to model this situation is to use the well-known one-factor Gaussian copula of Li (2000), in which each loss is interpreted as the time of default of a firm. Alternatively, we may use a multivariate Erlang mixture as follows.

$$F_\gamma(\mathbf{x}) = \gamma F_{QU}(\mathbf{x}) + (1-\gamma)\, F_I(\mathbf{x}), \tag{4.7}$$

where $F_I(x_1, \ldots, x_k) = \prod_{j=1}^{k} F_j(x_j)$ is the independent joint cdf. Distribution (4.7) preserves all the marginals and permits the correlation ranging from 0 ($\gamma = 0$) to the maximal correlations allowed within the subclass ($\gamma = 1$). In addition to the appealing features of Erlang mixtures, this model allows for the marginals being inputed directly, which avoids the cdf matching as did in the Gaussian copula model.

When $k = 2$, we may extend the above idea further by considering the following three mixture models:

$$F_{\gamma_1}(x,y) = \gamma_1 F_{QU}(x,y) + (1-\gamma_1)\, F_I(x,y),\ 0 \le \gamma_1 \le 1,$$
$$F_{\gamma_2}(x,y) = \gamma_2 F_{QL}(x,y) + (1-\gamma_2)\, F_I(x,y),\ 0 \le \gamma_2 \le 1,$$

and

$$F_{\gamma_{12}}(x,y) = \gamma_1 F_{QU}(x,y) + \gamma_2 F_{QL}(x,y) + (1-\gamma_1-\gamma_2)\, F_I(x,y),$$
$$\gamma_1 \ge 0,\ \gamma_2 \ge 0,\ \gamma_1 + \gamma_2 \le 1.$$

5. Aggregate and Excess Losses

Assume that there are k blocks of business or k types of policies. Let X_j, $j = 1, \ldots, k$, represent the associated individual losses. In order to model the aggregate loss arising from the portfolio in a tractable way, we usually assume the independence among individual losses in actuarial science. This assumption becomes unnecessary if we use a multivariate Erlang mixture since, (i) the aggregate and excess losses have again an Erlang mixture as shown below; and (ii) there is an efficient statistical algorithm that can fit a multivariate Erlang mixture to loss data well, as seen in the next section.

Throughout this section, we assume the individual losses X_j, $j = 1, \ldots, k$, have joint density (2.2). The following theorem identifies the distribution of the aggregate loss.

Theorem 5.1. *The aggregate loss $S_k = X_1 + \ldots + X_k$ has a univariate Erlang mixture with the mixing weights being the coefficients of the power series $P_{\mathbf{N}}(z, \ldots, z)$: for $i = 1, 2, \ldots$,*

$$\alpha_i^S = \sum_{m_1 + \cdots + m_k = i} \alpha_{\mathbf{m}}. \tag{5.1}$$

Proof. It is obvious from the proof of Theorem 2.2 that $S_k = \sum_{i=1}^{N_1 + \cdots + N_k} E_i$, where E_i's are iid exponential random variables with mean θ. Since $P_{\mathbf{N}}(z, \ldots, z)$ is the probability generating function of $N_1 + \ldots + N_k$, the mixing weights are its coefficients. □

Since the distribution of the aggregate loss is a univariate Erlang mixture, the mean, variance and other moments of S_k can be obtained easily. Furthermore, the associated risk measures such as value-at-risk (VaR) and Tail VaR (TVaR) of the aggregate loss S_k can be obtained explicitly as shown below.

Corollary 5.1. *The value-at-risk at confidence level p, $V = VaR_p(S_k)$, is the solution of equation*

$$e^{-V/\theta} \sum_{i=0}^{\infty} Q_i \frac{V^i}{\theta^i i!} = 1 - p \tag{5.2}$$

where $Q_i = \sum_{j=i+1}^{\infty} \alpha_j^S$ and α_j^S is given in (5.1).

The tail VaR at confidence level p, $TVaR_p(S_k)$, is given by

$$TVaR_p(S_k) = \frac{\theta e^{-V/\theta}}{1-p} \sum_{i=0}^{\infty} Q_i^* \frac{V^i}{\theta^i i!} + V, \tag{5.3}$$

where $Q_i^* = \sum_{j=i}^{\infty} Q_j = \sum_{j=i+1}^{\infty} (j-i)\, \alpha_j^S$.

Proof. They follows from 5.1 and the properties of a univariate Erlang mixture (see, for example, Willmot and Woo (2007) or Lee and Lin (2010)). □

The stop-loss premium of S_k at deductible level $d, E\{(S_k - d)_+\}$, where $x_+ = \max(x, 0)$, is obtainable immediately from (5.3):

$$E\{(S_k - d)_+\} = \theta e^{-d/\theta} \sum_{i=0}^{\infty} Q_i^* \frac{d^i}{\theta^i i!}. \tag{5.4}$$

In the following, we give the sharp upper bounds for the variance and the stop-loss premiums of S_k with the help of quasi-comonotonicity. To do that, we begin with a result in convex ordering. A random variable X is said to be smaller than a random variable Y in convex ordering, if for any convex function $h(x)$, $E\{h(X)\} \leq E\{h(Y)\}$. See p. 55 of Shaked and Shanthikumar (1994).

Theorem 5.2. *For the aggregate loss $S_k = X_1 + \ldots + X_k$, let $S_k^U = X_1^U + \ldots + X_k^U$ be the corresponding quasi-comonotonic sum. Then, S_k is smaller than S_k^U in convex ordering:*

$$S_k \leq_{cx} S_k^U. \tag{5.5}$$

Proof. From the proof of Theorem 5.1, we may write $S_k = \sum_{i=1}^{N_1+\cdots+N_k} E_i$, where E_i's are iid exponential random variables. Similarly, $S_k^U = \sum_{i=1}^{N_1^U+\cdots+N_k^U} E_i$. Theorem 7 of Dhaene et al. (2002a) means

$$N_1 + \ldots + N_k \leq_{cx} N_1^U + \ldots + N_k^U.$$

It follows from Theorem 2.A.7 of Shaked and Shanthikumar (1994) that

$$S_k \leq_{cx} S_k^U.$$

□

Theorem 5.2 enables us to obtain the sharp upper bounds for the variance and the stop-loss premium as below.

Corollary 5.2.

$$Var(S_k) \leq Var(S_k^U) \tag{5.6}$$

Moreover, for any deductible d, we have

$$E\{(S_k - d)_+\} \leq E\{(S_k^U - d)_+\}. \tag{5.7}$$

Proof. These are the direct results of Formulas (2.A.4) and (2.A.5) of Shaked and Shanthikumar (1994). □

We now turn to the distribution of the multivariate excess losses that is important for the calculation of economic capital for an insurance portfolio.

Let $\mathbf{d} = (d_1, \ldots, d_k)$ be deductible levels (or economic capitals) of the individual losses $\mathbf{X} = (X_1, \ldots, X_k)$ from an insurance portfolio. The associated multivariate excess losses may thus be defined as the conditional random vector $\mathbf{Y_d} = \mathbf{X} - \mathbf{d} \,|\, \mathbf{X} > \mathbf{d}$. We identify the distribution of $\mathbf{Y_d}$ in the next theorem, when $\mathbf{X}$ has a multivariate Erlang mixture.

Theorem 5.3. *When* $\mathbf{X}$ *has a multivariate Erlang mixture, the joint density of* $\mathbf{Y_d}$ *is again a multivariate Erlang mixture with the same scale parameter. Its mixing weights are given by (5.9) below.*

Proof. Let $f_{\mathbf{d}}(\mathbf{y})$ be the joint distribution of $\mathbf{Y_d}$. Then

$$f_{\mathbf{d}}(\mathbf{y}) = \frac{f(\mathbf{y}+\mathbf{d})}{\bar{F}(\mathbf{d})}.$$

Since

$$p(y+d;m,\theta) = e^{-d/\theta} \sum_{i=1}^{m} \frac{(d/\theta)^{m-i}}{(m-i)!}\, p(y;i,\theta),$$

$$\begin{aligned}
f_{\mathbf{d}}(\mathbf{y}) &= \frac{1}{\bar{F}(\mathbf{d})}\, f(\mathbf{y}+\mathbf{d}) \\
&= \frac{e^{-(\sum_{j=1}^{k} d_j)/\theta}}{\bar{F}(\mathbf{d})} \sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \alpha_{\mathbf{m}} \prod_{j=1}^{k} \sum_{i_j=1}^{m_j} \frac{(d/\theta)^{m_j-i_j}}{(m_j-i_j)!}\, p(y_j;\, i_j,\theta) \\
&= \frac{e^{-(\sum_{j=1}^{k} d_j)/\theta}}{\bar{F}(\mathbf{d})} \sum_{i_1=1}^{\infty} \cdots \sum_{i_k=1}^{\infty} \sum_{m_1=i_1}^{\infty} \cdots \sum_{m_k=i_k}^{\infty} \alpha_{\mathbf{m}} \prod_{j=1}^{k} \frac{(d/\theta)^{m_j-i_j}}{(m_j-i_j)!}\, p(y_j;\, i_j,\theta) \\
&= \sum_{i_1=1}^{\infty} \cdots \sum_{i_k=1}^{\infty} c_{\mathbf{i}} \prod_{j=1}^{k} p(y_j;\, i_j,\theta),
\end{aligned} \tag{5.8}$$

where $\mathbf{i} = (i_1, \ldots, i_k)$ and

$$c_{\mathbf{i}} = \frac{e^{-(\sum_{j=1}^{k} d_j)/\theta}}{\bar{F}(\mathbf{d})} \sum_{m_1=i_1}^{\infty} \cdots \sum_{m_k=i_k}^{\infty} \alpha_{\mathbf{m}} \prod_{j=1}^{k} \frac{(d/\theta)^{m_j-i_j}}{(m_j-i_j)!}. \tag{5.9}$$

Note that $\bar{F}(\mathbf{d})$ is obtained in the second line of (4.2). □

We remark that Theorem 5.3 implies that the use of a multivariate Erlang mixture may have certain advantages over the use of a copula for risk management. In general, the joint distribution of multivariate excess losses can not be

https://doi.org/10.2143/AST.42.1.2160739 Published online by Cambridge University Press

obtained explicitly using a copula approach. As a result, risk measures related to an insurance/investment portfolio are often obtained by simulation. Second, when we apply a multivariate model to the times of default for a collection of firms, it is extremely important to know how the correlations among the times of default change over time. Since the explicit joint conditional distribution of the times of default, conditioning on a future time point (say one day later), is not available when using a copula, there is no meaningful connection between the original model and a re-calibrated model at the future time. On the other hand, we may use a multivariate Erlang mixture such that the random variable X_j represents the default time of Firm j. By setting $d_1 = \ldots = d_k = t$ that represents the future time for model evaluation, we could then easily compare the mixing weights of the joint conditional distribution from the original model using (5.8) and a re-calibrated model at the future time.

6. Parameter Estimation: an EM Algorithm

We have shown in the previous sections that the class of multivariate Erlang mixtures has many desirable properties. We have also shown that there is a simple method in the proof of Theorem 2.1 that can approximate any positive continuous multivariate distribution. However, if we employ this simple method when treating the data as the empirical distribution, the parameters in the mixture do not maximize the likelihood in general. Improving the accuracy or likelihood by increasing the number of component distributions, or equivalently lowering the value of θ is undesirable, as it often risks the problem of over-fitting in many situations. Hence, there is a need to find an efficient fitting tool, otherwise the class might not be as useful in practice. In this section, we present an EM algorithm that can efficiently fit a multivariate Erlang mixture to multivariate positive data. Our goal is to optimize the parameters by applying the EM algorithm and to reduce the number of parameters needed for fitting. The EM algorithm proposed in this section is a simple extension of that in Lee and Lin (2010) for the univariate case.

6.1. The EM algorithm

The EM algorithm was proposed in Dempster et al. (1977). It is an iterative algorithm for finding the maximum likelihood estimate of the parameters of an underlying distribution from a set of incomplete data and is particularly useful in estimating the parameters of a finite mixture.

Consider now a data set of size n: $\mathbf{x}_v = (x_{1v}, x_{2v}, \ldots, x_{kv})$, $v = 1, \ldots, n$. We are to use a k-variate finite Erlang mixture to fit the data. The set of shape parameters of the Erlang distributions are preset and we denote it by $\mathcal{M}$, i.e. $\mathcal{M}$ contains a finite number of $\mathbf{m}$'s. For notational convenience, hereafter if $\mathbf{m} \notin \mathcal{M}$, we set $\alpha_{\mathbf{m}} = 0$. In this case, the set of parameters (denoted by Φ) to be estimated are the scale parameter θ and the weights $\alpha_{\mathbf{m}}$ where $\mathbf{m} \in \mathcal{M}$.

Assuming that the parameter values from the $l-1^{st}$ iteration are $\Phi^{(l-1)}$, the posterior probability function given the observation $\mathbf{x}_v$ and $\Phi^{(l-1)}$ in the E-step is, for $\mathbf{m} \in \mathcal{M}$,

$$q\left(\mathbf{m} \,|\, \mathbf{x}_v, \Phi^{(l-1)}\right) = \frac{\alpha_{\mathbf{m}}^{(l-1)} \prod_{j=1}^{k} p(x_{jv}, m_j, \theta)}{\sum_{r_1=1}^{\infty} \cdots \sum_{r_k=1}^{\infty} \alpha_{\mathbf{r}}^{(l-1)} \prod_{j=1}^{k} p(x_{jv}, r_j, \theta)} \tag{6.1}$$

and the corresponding expected log-likelihood,

$$Q(\Phi \,|\, \Phi^{(l-1)}) = \sum_{v=1}^{n} \sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \left(\ln \alpha_{\mathbf{m}} - \frac{1}{\theta} \sum_{j=1}^{k} x_{jv} - \ln \theta \sum_{j=1}^{k} m_j \right) q\left(\mathbf{m} \,|\, \mathbf{x}_v, \Phi^{(l-1)}\right). \tag{6.2}$$

The M-step is then to maximize the expectation:

$$\Phi^{(l)} = \max_{\Phi} Q(\Phi \,|\, \Phi^{(l-1)}) \tag{6.3}$$

Taking partial derivatives of $Q(\Phi|\Phi^{(l-1)})$ with respect to the parameters and with the constraint $\sum_{\mathbf{m} \in \mathcal{M}} \alpha_{\mathbf{m}}^{(l)} = 1$ yields

$$\alpha_{\mathbf{m}}^{(l)} = \frac{1}{n} \sum_{v=1}^{n} q\left(\mathbf{m} \,|\, \mathbf{x}_v, \Phi^{(l-1)}\right), \ \ \mathbf{m} \in \mathcal{M}, \tag{6.4}$$

and

$$\theta^{(l)} = \frac{\sum_{v=1}^{n} \sum_{j=1}^{k} x_{jv}}{n \sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \left(\sum_{j=1}^{k} m_j \right) \alpha_{\mathbf{m}}^{(l)}} \tag{6.5}$$

The iterations will continue until the convergence of the expected log-likelihood to a predefined tolerance is reached.

Good initial estimates of the parameters are crucial for the EM algorithm. It helps significantly reduce the computing time by requiring fewer iterations. It is especially important when dealing with multivariate data as the number of parameters to optimize increase nonlinearly. In Lee and Lin (2010), we propose a method to determine the initial value of θ and the weights of a univariate Erlang mixture, and a model selection procedure using Schwarz's Bayesian Information Criterion (BIC) that penalizes over-fitting. The method and the procedure can be easily adopted for the multivariate case and we omit the details.

6.2. Fitting Data from a Multivariate Lognormal Distribution

In this subsection, we fit data generated from a multivariate lognormal distribution of 12 dimensions to demonstrate the efficiency of the EM algorithm for high-dimensional data.

Let $Z_j, j = 1, 2, \ldots, 12$, be iid lognormal random variables with parameters μ and σ, and

$$X_i = \prod_{j=1}^{i} Z_j, \; i = 1, 2, \ldots, 12, \tag{6.6}$$

Then, $(X_1, \ldots, X_{12})$ has a multivariate lognormal distribution. This model is motivated by the applications in the pricing of options and equity-indexed annuities (EIA). Consider the price of a risky asset or an equity index that follows a geometric Brownian motion with drift 12μ and volatility 12σ over a one-year period. Thus, $X_1, \ldots, X_{12}$ represent the prices of the asset at the end of each month. If we could model the joint distribution of $(X_1, \ldots, X_{12})$ using a multivariate Erlang mixture, the distribution of the sum $S_{12} = \sum_{i=1}^{12} X_i$ is readily obtained (Theorem 5.1). Note that $\frac{1}{12}(S_{12} - 12K)_+$ is the payoff of an arithmetic Asian option with strike price K maturing at the end of one year, it follows from (5.4) that the price of this Asian option has a closed-form solution. Furthermore, $\frac{1}{12}(S_{12} - 12K)_+$ may also be viewed as the payoff of the so-called Asian-end minimum guarantee with monthly averaging of an EIA, where K is the minimum guarantee required by the non-forfeiture law in the United States. See Lin and Tan (2003).

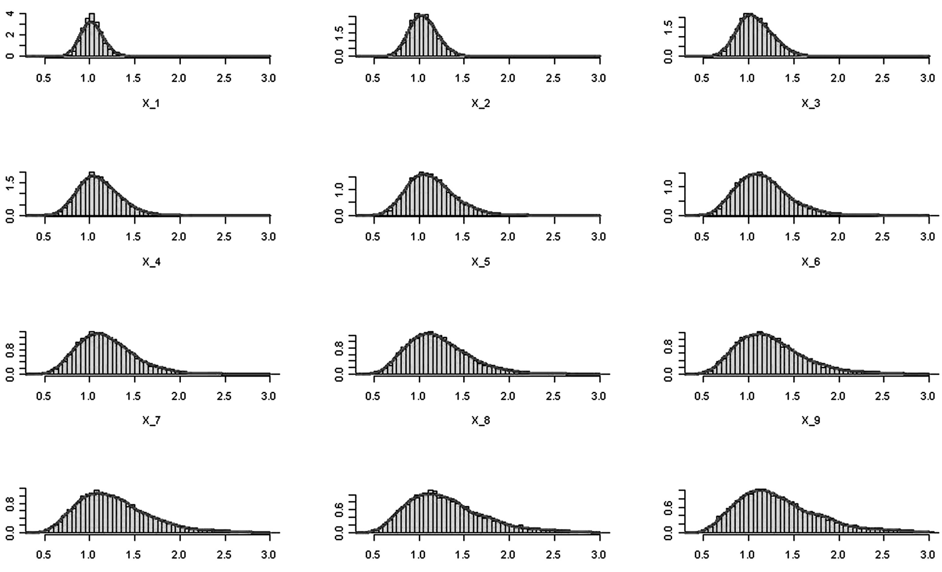

FIGURE 6.1: Histograms of 12 marginal lognormal distributions and marginal densities of fitted distribution.

https://doi.org/10.2143/AST.42.1.2160739 Published online by Cambridge University Press

We now assume that $\mu = 2.5\%$ and $\sigma = 10\%$ and simulate 8000 observations from $(X_1, X_2, \ldots, X_{12})$. The estimated values of the parameters are given in Table B.1 of Appendix B. Figure 6.1 shows the fitting results for all the 12 marginals.

However, the validity of using the marginals to represent the fitness of the model is questionable as the dependence structure is not shown in these plots. To address the issue and from a viewpoint of applications to EIA, we investigate the fitness of the density of S_{12} obtained by using the formula in Theorem 5.1. Since a poor overall fitting to the multivariate data would in general result in a poor fitting to the aggregated data, fitting to the aggregated data could be a good measure for the goodness of fit. Figure 6.2 provides the visual fitting result in this regard. It shows the histogram of the aggregated data and the density represented by the solid line.

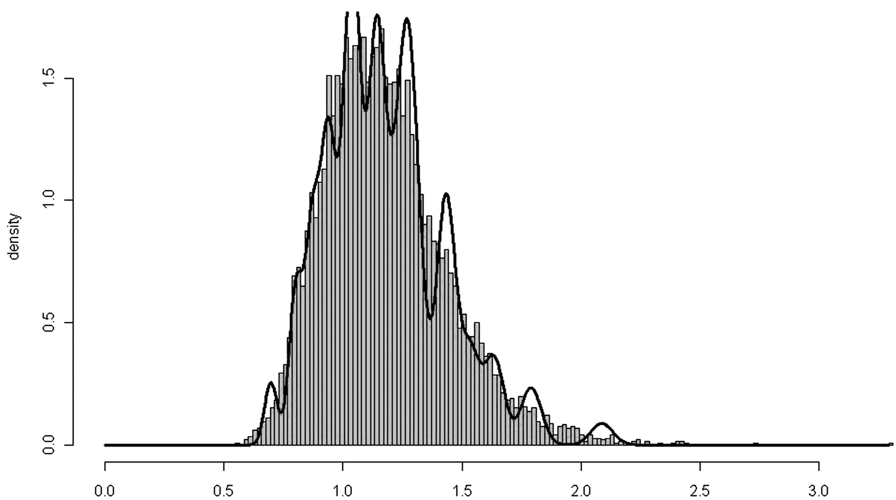

FIGURE 2: Histogram of aggregated data and density of sum of marginals from fitted distribution.

We can also check the fitness quantitatively by performing several common statistical tests. The tests we use in this section are the Chi-square test, the Kolmogorov-Smirnov test and the Anderson Darling test.

TABLE 6.3

STATISTICAL TESTS FOR FITNESS OF ERLANG MIXTURE TO AGGREGATED DATA

Test	Statistic	p-value	Accepted at 5% significant level?
Chi Square Test	818.32	0.3099	Yes
K-S Test	0.05	0.27	Yes
AD Test	0.4378	0.2228	Yes

The fitted distribution passed all three tests with significant margins. It implies that the fitted distribution is a good representation of the aggregate distribution. Another quantitative aspect is to compare the first 5 raw moments of the empirical and fitted distributions:

TABLE 6.4

FIRST 5 MOMENTS OF EMPIRICAL AND FITTED DISTRIBUTIONS

Moment	Empirical Distribution	Fitted Distribution	Fitted/ Empirical	Percentage Difference (%)
1	1.1791	1.1791	1.00000	0.0000%
2	1.4566	1.4588	0.9985	0.1511%
3	1.8871	1.8971	0.9947	0.5284%
4	2.5654	2.5985	0.9829	1.2712%
5	3.6605	3.7592	0.9737	2.6237%

Recall that higher order moments amplify the deviations at the upper tail. The percentage difference of the 3rd order moment is 0.5284% which is negligible. Again, it shows that the fitted distribution is a good fit to the data.

We have also performed other numerical examples to examine the fitness of the multivariate Erlang mixture to data. In particular, we have used data generated from the Gumbel-Hougaard copula and Frank copula with fixed Pareto marginals to examine the fitness. All the fitting results we have obtained are very satisfactory. To avoid repetitiveness, they are not presented in this paper.

6.3. Construction of bivariate Erlang mixtures with extreme dependence

In this subsection, we construct two bivariate Erlang mixtures from two given marginal distributions. One has Spearman's rho very close to 1 and the other has Spearman's rho very close to –1. The purpose of the construction is twofold: (i) to show that the multivariate Erlang mixture is indeed flexible in terms of dependence structure even in an extreme case; and (ii) to demonstrate that the EM algorithm in Section 6.1 not only is an efficient fitting algorithm but also a useful tool in providing a multivariate distribution with a given level of dependence.

We begin with two gamma marginal distributions: one has $\alpha_1 = 1.7$ and $\theta_1 = 1.67$, and the other has $\alpha_2 = 2.6$ and $\theta_2 = 2$. Note that both are not Erlang distributions. The Fréchet-Hoeffding upper bound of these two distributions gives a bivariate distribution with Spearman's $\rho = 1$ and the Fréchet-Hoeffding lower bound of them gives a bivariate distribution with Spearman's $\rho = -1$. We generate data from both bivariate distributions, respectively, and then apply the EM algorithm in Section 6.1 to each.

In the case of $\rho = 1$, we obtained a bivariate Erlang mixture with 19 components. The parameters of the mixture are given in in the left three columns

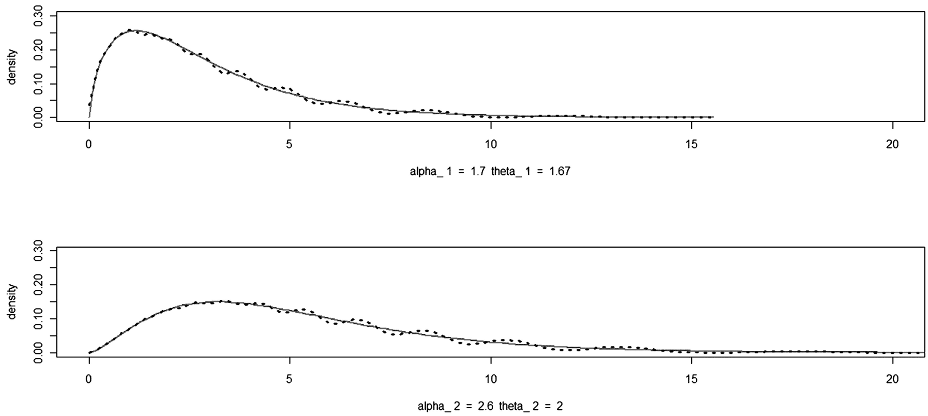

FIGURE 6.3: Gamma densities (solid lines) versus marginal densities of fitted distribution (dotted lines).

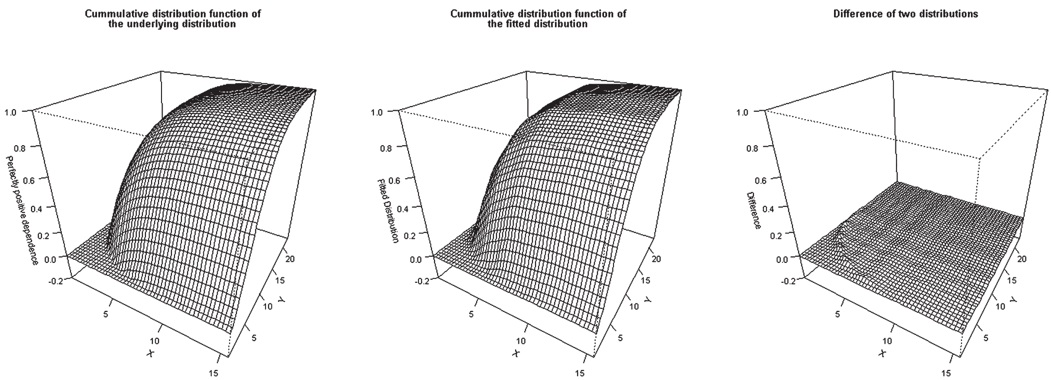

FIGURE 6.4: Comparison of actual joint cdf to fitted joint cdf.

of Table B.2 in Appendix B. The marginal and joint distributions are plotted in Figures 6.3 and 6.4.

Spearman's rho of the fitted distribution is 0.9733. The value of ρ will be higher if we further increase the number of components. The fitting results are good for both the marginals and the joint distribution.

In the case of $\rho = -1$, we have a bivariate Erlang mixture of 12 components. The parameters of the mixture are given in the last three columns of Table B.2 in Appendix B. The marginal and joint distributions are plotted in Figures 6.5 and 6.6 below.

Spearman's rho of the fitted distribution is -0.9728. The fitting results are again good for both the marginals and the joint distribution. Also note that since 12 components are used, the fitting results are slightly worse than those in Figures 6.3 and 6.4.

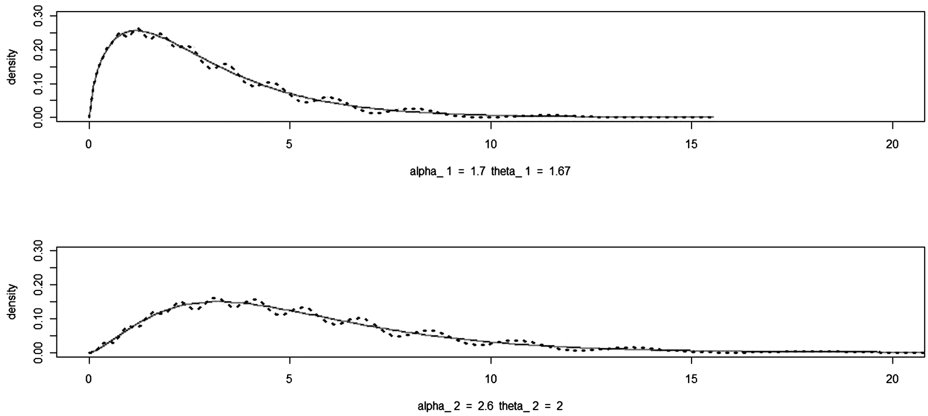

FIGURE 6.5: Gamma densities (solid lines) versus marginal densities of fitted distribution (dotted lines).

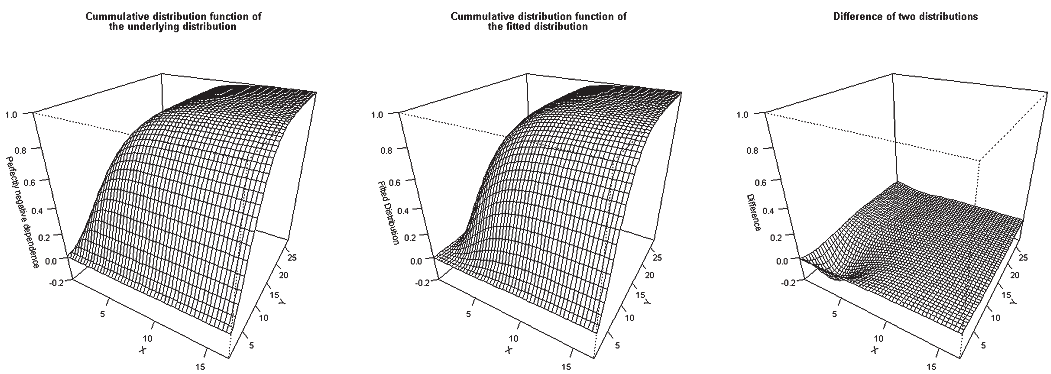

FIGURE 6.6: Comparison of actual joint cdf to fitted joint cdf.

7. Concluding Remarks and Discussion

In this paper, we introduced a class of multivariate Erlang mixtures. We have shown that the class has many desirable properties and could be useful for multivariate modeling in insurance and risk management. An efficient EM algorithm is developed to fit a multivariate Erlang mixture to multivariate positive data.

Possible applications of the multivariate Erlang mixture include pricing path dependent options and valuation of equity indexed annuities (EIAs). It is a common practice in option pricing to assume that the value of the underlying asset of an option follows a geometric Brownian motion or more generally the logarithm of a Gaussian process. As discussed in Subsection 6.2, when the value of an asset follows such a process, the price of an arithmetic Asian option written on the asset under the constant interest rate does not have a

https://doi.org/10.2143/AST.42.1.2160739 Published online by Cambridge University Press

closed form because the sum of lognormal random variables is no longer lognormal. However, if one uses a multivariate Erlang mixture to approximate the joint distribution of the lognormal random variables, the distribution of the sum is a univariate Erlang mixture. The latter enables us to obtain a closed form for the price of the arithmetic Asian option. Furthermore, if a random vector has a multivariate Erlang mixture, their maximum has a univariate Erlang mixture. As a result, the price of the usual lookback options has a closed form when we use the same approximation. These results are particularly useful in the valuation of EIAs when the guarantee of an EIA is linked to the average of the index or is of high-water mark. Another possible application is the valuation of basket options. Again, if we assume the values of the stocks in a portfolio at maturity have a multivariate Erlang mixture, the total value of the portfolio has a univariate Erlang mixture and the price of a basket option may be obtained explicitly. We are currently working on these problems and intend to compare our numerical results with those in Dhaene et al. (2002b) and Vanduffel et al. (2005) in which the authors propose comonotonicity based bounds to estimate the stop-loss premium of the sum of lognormal random variables and hence the price of arithmetic Asian options.

The class of multivariate Erlang mixtures can be expanded by not restricting $\mathbf{N}$ to be a positive counting random vector. That is, each component N_j is a usual counting random variable, taking all the non-negative integers, which in turn allows for multivariate Erlang mixtures to have a probability mass at zero. For example, we may let $\mathbf{N}$ be a multivariate Poisson random vector with common shock such that $N_j = J_0 + J_j$, $j = 1, 2, \ldots$, where $J_0, J_1, \ldots, J_k$ are mutually independent Poisson random variables with means $\lambda_0, \lambda_1, \ldots, \lambda_k$, respectively. See Johnson et al. (1997), p. 139. In this case, it follows from (2.4) that the moment generating function is

$$\begin{aligned} M_{\mathbf{X}}(\mathbf{z}) &= P_{\mathbf{N}}\left(\frac{1}{1-\theta z_1}, \cdots, \frac{1}{1-\theta z_k}\right) \\ &= \exp\left\{\frac{\lambda_0}{\prod_{j=1}^{k}(1-\theta z_j)} + \sum_{j=1}^{k}\frac{\lambda_j}{(1-\theta z_j)} - \sum_{j=0}^{k}\lambda_j\right\}. \end{aligned}$$

The above multivariate compound Poisson distribution belongs to the multivariate Tweedie family with dispersion parameter $p = 1.5$. For the multivariate Tweedie family and its properties and applications in actuarial science, see Furman and Landsman (2010) and references therein. We caution that the multivariate Tweedie family is not dense in the space of all non-negative multivariate distributions.

Note that the recursive formulas (6.4) and (6.5) are no longer applicable when we expand the class to include non-negative counting random vectors. In principle, the standard EM algorithm will still work but an optimization technique needs to be used for the M-step. Alternatively we may use a naïve approach by partitioning a data set such that each sub-dataset $\mathtt{D_s}$ is a subset of the subspace $\{\mathbf{x}; x_i = 0, i \notin \mathtt{s}\}$, where $\mathtt{s} \in \mathcal{S}$ and $\mathcal{S}$ is the collection of all

https://doi.org/10.2143/AST.42.1.2160739 Published online by Cambridge University Press

the subsets of $\{1, 2, \ldots, k\}$, including the null set. The EM algorithm (6.4) and (6.5) are then applied to estimate the Erlang mixture $F_{\mathtt{s}}(\mathbf{x})$ on the subspace $\{\mathbf{x};\ x_i = 0,\ i \notin \mathtt{s}\}$, using the positive data taken from $\mathtt{D}_{\mathtt{s}}$. The final estimated Erlang mixture will then be

$$\sum_{s \in \mathcal{S}} \beta_{\mathtt{s}} F_{\mathtt{s}}(\mathbf{x}),$$

where $\beta_{\mathtt{s}}$ is the ratio of the size of $\mathtt{D}_{\mathtt{s}}$ to the size of the entire data set. A shortcoming of this approach might be that there could be too many parameters. Seeking more effective algorithms for estimation of extended multivariate Erlang mixtures would be an interesting project for future research.

Appendix A: Proof of Theorem 2.1

The proof is very similar to that in Lee and Lin (2010). We first show in the following that for a k-variate distribution on the set $\mathbb{R}_+^k = \{\mathbf{x} = (x_1, \ldots, x_k);\ x_j > 0,\ j = 1, \ldots, k\}$ with density $f(\mathbf{x})$, there is a sequence of k-variate Erlang mixtures such that their characteristic functions converge pointwise to the characteristic function of $f(\mathbf{x})$.

For any given value θ, consider a k-variate Erlang mixture with the following density

$$\hat{f}(\mathbf{x}|\theta) = \sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \alpha_{\mathbf{m}} \prod_{j=1}^{k} p(x_j, m_j, \theta). \tag{A.1}$$

where

$$\alpha_{\mathbf{m}} = \int_{(m_1-1)\theta}^{m_1\theta} \cdots \int_{(m_k-1)\theta}^{m_k\theta} f(\mathbf{x})\, dx_k \cdots dx_1. \tag{A.2}$$

Let $\varphi(\mathbf{z}) = \int_0^\infty \cdots \int_0^\infty e^{\mathbf{i}\mathbf{z}\cdot\mathbf{x}} f(\mathbf{x})\, d\mathbf{x}$ be the characteristic function of $f(\mathbf{x})$, where $\mathbf{i} = \sqrt{-1}$, $\mathbf{z} = (z_1, \ldots, z_k)$, $d\mathbf{x} = dx_k \ldots dx_1$ and $\mathbf{z}\cdot\mathbf{x}$ is the usual inner product. Similarly, $\varphi_\theta(\mathbf{z})$ is the characteristic function of $\hat{f}(\mathbf{x}|\theta)$.

From (A.2) and the fact that the characteristic function of $p(x; m, \theta)$ is $(1 - \mathbf{i}\theta z)^{-m}$, we have

$$\begin{aligned}
\varphi_\theta(\mathbf{z}) &= \sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \alpha_{\mathbf{m}} \prod_{j=1}^{k} (1 - \mathbf{i}\theta z_j)^{-m_j} \\
&= \sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \int_{(m_1-1)\theta}^{m_1\theta} \cdots \int_{(m_k-1)\theta}^{m_k\theta} \left[\prod_{j=1}^{k} (1 - \mathbf{i}\theta z_j)^{-m_j} \right] f(\mathbf{x})\, d\mathbf{x} \\
&= \sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \int_{(m_1-1)\theta}^{m_1\theta} \cdots \int_{(m_k-1)\theta}^{m_k\theta} \left[\prod_{j=1}^{k} (1 - \mathbf{i}\theta z_j)^{-\lceil x_j/\theta \rceil} \right] f(\mathbf{x})\, d\mathbf{x} \\
&= \int_0^\infty \cdots \int_0^\infty \left[\prod_{j=1}^{k} (1 - \mathbf{i}\theta z_j)^{-\lceil x_j/\theta \rceil} \right] f(\mathbf{x})\, d\mathbf{x},
\end{aligned} \tag{A.3}$$

where the ceiling function $\lceil x \rceil$ gives the smallest integer greater than or equal to x. Since $(1 - \mathbf{i}\theta z)^{-\lceil x/\theta \rceil}$ is bounded from above as long as $\theta|z| < 1$ and $\lim_{\theta \to 0}(1 - \mathbf{i}\theta z)^{-\lceil x/\theta \rceil} = e^{\mathbf{i}zx}$, $\lim_{\theta \to 0} \varphi_\theta(\mathbf{z}) = \varphi(\mathbf{z})$ for all $\mathbf{z}$ by the Dominance Convergence Theorem. By Levy's Continuity Theorem on convergence in distribution (Billingsley, 1995, pp. 381-3), the joint distribution function of (A.1) converges to the joint distribution function of $f(\mathbf{x})$ pointwise, as $\theta \to 0$.

Appendix B: Estimated parameter values for fitted distributions

TABLE B.1.

Shape parameters and estimated weights of fitted distribution with $\theta = 0.01253039$

	m_{i_1}	m_{i_2}	m_{i_3}	m_{i_4}	m_{i_5}	m_{i_6}	m_{i_7}	m_{i_8}	m_{i_9}	$m_{i_{10}}$	$m_{i_{11}}$	$m_{i_{12}}$	$\alpha_{\mathbf{m}}$
1	75	70	65	62	59	57	55	54	53	52	52	52	0.03519954
2	77	75	73	72	73	75	78	82	86	91	97	101	0.06750167
3	75	70	66	64	63	63	64	65	68	70	73	75	0.05352882
4	80	79	79	81	83	86	91	98	106	115	122	129	0.06488830
5	80	81	84	89	96	103	109	113	114	114	112	112	0.06019880
6	83	86	90	94	99	105	111	120	129	138	145	150	0.08021910
7	80	79	78	77	75	72	69	66	64	62	61	61	0.06330692
8	79	78	77	77	77	77	77	77	77	78	79	80	0.11508296
9	82	83	84	86	87	89	91	92	94	94	95	97	0.13055435
10	85	88	94	100	109	119	129	143	158	171	182	191	0.03218294
11	89	99	109	116	125	133	139	143	146	149	152	156	0.04549171
12	85	89	92	93	92	90	87	83	79	77	76	76	0.05818215
13	87	92	97	99	100	100	100	102	105	110	116	121	0.06408133
14	87	93	99	103	105	106	105	102	99	96	93	93	0.05392744
15	88	96	104	112	119	122	123	123	122	122	121	122	0.05431533
16	91	103	114	128	141	156	167	178	189	199	209	214	0.02133865

TABLE B.2.

shape parameters and mixing weights of fitted distributions

	$\rho = 1$			$\rho = -1$		
	$\theta = 0.04178421$			$\theta = 0.04139381$		
	m_{i_1}	m_{i_2}	$\alpha_{\mathbf{m}}$	m_{i_1}	m_{i_2}	$\alpha_{\mathbf{m}}$
1	1	6	0.008375923	2	436	0.008375923
2	2	8	0.028893781	6	327	0.028893781
3	3	12	0.059347992	12	258	0.059347992

	$\rho = 1$			$\rho = -1$		
	$\theta = 0.04178421$			$\theta = 0.04139381$		
	m_{i_1}	m_{i_2}	$\alpha_{\mathbf{m}}$	m_{i_1}	m_{i_2}	$\alpha_{\mathbf{m}}$
4	6	20	0.096291043	20	206	0.096291043
5	9	28	0.130629033	31	164	0.130629033
6	12	34	0.150767902	45	130	0.150767902
7	13	36	0.154488303	62	101	0.154488303
8	17	44	0.139608365	84	77	0.139608365
9	20	50	0.108410072	111	56	0.108410072
10	27	62	0.072483792	146	39	0.072483792
11	28	65	0.038706564	195	24	0.038706564
12	39	83	0.011997230	277	11	0.011997230
13	52	104	0.072483792			
14	69	130	0.038706564			
15	91	162	0.011997230			
16	118	201	0.011997230			
17	153	250	0.072483792			
18	202	316	0.038706564			
19	282	422	0.011997230			

Acknowledgments

This research was partly supported by grants from the Natural Sciences and Engineering Council of Canada (NSERC) and the Mathematics of Information Technology and Complex Systems NCE (MITACS). The authors are grateful to Jan Dhaene, Gordon Willmot and the two referees for their valuable comments.

References

Asmussen, S. and Albrecher, H. (2010) *Ruin Probabilities*, 2nd Edition, Singapore: World Scientific Publishing.

Asmussen, S., Nerman, O. and Olsson, M. (1996) "Fitting phase-type distributions via the EM algorithm", *Scandinavian Journal of Statistics* **23(4)**, 419-441.

Assaf, D., Langberg, N.A., Savits, T.H. and Shaked, M. (1984) "Multivariate phase-type distributions", *Operations Research* **32(3)**, 688-702.

Billingsley, P. (1995) *Probability and Measure*, Third Edition, John Wiley and Sons, Hoboken, NJ.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society*, *Series B (Methodological)*, **39(1)**, 1-38.

Denuit, M., Dhaene, J., Goovaerts, M.J. and Kaas, R. (2005) *Actuarial Theory for Dependent Risks*: *Measures*, *Orders and Models.* Wiley and Sons.

Dhaene, J., Denuit, M., Goovaerts, M.J., Kaas, R. and Vyncke, D. (2002) "The concept of comonotonicity in actuarial science and finance: theory", *Insurance*: *Mathematics and Economics*, **31**, 3-33.

Dhaene, J., Denuit, M., Goovaerts, M.J., Kaas, R. and Vyncke, D. (2002) "The concept of comonotonicity in actuarial science and finance: applications", *Insurance*: *Mathematics and Economics*, **31**, 133-161.

Dhaene, J., Vanduffel, S., Tang, Q., Goovaerts, M.J., Kaas, R. and Vyncke, D. (2006) Risk measures and comonotonicity: a review. *Stochastic Models*, **22**, 573-606.

Eisele, K-T. (2005) "EM algorithm for bivariate phase distributions", ASTIN Colloquium, Zurich, Switzerland. http://www.actuaries.org/ASTIN/Colloquia/Zurich/Eisele.pdf

Fréchet, M.R. (1951) "Sur les tableaux de corrélation dont les marges sont données", *Annal University Lyon Science*, **4**, 53-84.

Fréchet, M.R. (1958) "Remarques au sujet de la note précédente", *C. R. Acad. Sci. Paris Sr. I Math.*, **246**, 2719-2720.

Frees, E.W. and Valdez, E.A. (1998) "Understanding relationships using copulas", *North American Actuarial Journal*, **2(1)**, 1-25.

Furman, E. and Landsman, Z. (2010) "Multivariate Tweedie distributions and some related capital-at-risk analyses", *Insurance*: *Mathematics and Economics*, **46**, 351-361.

Genest, C., Gerber, H.U., Goovaerts, M.J. and Laeven, R.J.A. (2009) "Editorial to the special issue on modeling and measurement of multivariate risk in insurance and finance". *Insurance*: *Mathematics and Economics*, **44**, 143-145.

Hoeffding, W. (1940) "Scale-invariant correlation theory, In N.I. Fisher and P.K. Sen (eds.), The Collected Works of Wassily Hoeffding", New York: Springer-Verlag.

Hoeffding, W. (1940) "Scale-invariant correlation measures for discontinuous distributions, In N.I. Fisher and P.K. Sen (eds.), The Collected Works of Wassily Hoeffding", New York: Springer-Verlag.

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

Johnson N.L., Kotz, S. and Balakrishnan, A.V. (1997) *Discrete Multivariate Distributions.* Wiley and Sons.

Li, D.X. (2000) "On default correlation: a copula approach", *Journal of Fixed Income*, **9**, March, 43-54.

Lee, C.K. and Lin, X.S. (2010) "Modeling and Evaluating Insurance Losses via Mixtures of Erlang Distributions", *North American Actuarial Journal*, **14(1)**, 107-130.

Lin, X.S. and Tan, K.S. (2003) "Valuation of equity-indexed annuities under stochastic interest rate", *North American Actuarial Journal*, **7(4)**, 72-91.

Mikosch, T. (2006) "Copulas: tales and facts", *Extremes*, **9**, 3-20.

Nelson, R.B. (1999) *An Introduction to Copulas.* Lecture Notes in Statistics 139, Springer, New York.

Shaked, M. and Shanthikumar, J.G. (1994) *Stochastic Orders and Their Applications.* Academic Press, San Diego.

Vanduffel, S., Hoedemakers, T. and Dhaene, J. (2005) "Comparing approximations for risk measures of sums of non-independent lognormal random variables", *North American Actuarial Journal*, **9(4)**, 71-82.

Willmot, G.E and Woo, J.K. (2007) "On the class of Erlang mixtures with risk theoretic applications", *North American Actuarial Journal*, **11(2)**, 99-115.

Simon C.K. Lee
Silen Trading and Consulting Inc.
E-Mail: cklee.simon@gmail.com

X. Sheldon Lin (corresponding author)
Department of Statistics
University of Toronto
Toronto, Ontario M5S 3G3
Canada
E-Mail: sheldon@utstat.utoronto.ca