Moment-Based Approximation with Mixed Erlang Distributions

by Hélène Cossette, David Landriault, Etienne Marceau,
and Khouzeima Moutanabbir

ABSTRACT

Moment-based approximations have been extensively analyzed over the years (see, e.g., Osogami and Harchol-Balter 2006 and references therein). A number of specific phase-type (and non phase-type) distributions have been considered to tackle the moment-matching problem (see, for instance, Johnson and Taaffe 1989). Motivated by the development of more flexible moment-based approximation methods, we develop and examine the use of finite mixture of Erlangs with a common rate parameter for the moment-matching problem. This is primarily motivated by Tijms (1994) who shows that this class of distributions can approximate any continuous positive distribution to an arbitrary level of accuracy, as well as the tractability of this class of distributions for various problems of interest in quantitative risk management. We consider separately situations where the rate parameter is either known or unknown. For the former case, a direct connection with a discrete moment-matching problem is established. A parallel to the s-convex stochastic order (e.g., Denuit et al. 1998) is also drawn. Numerical examples are considered throughout.

1. Introduction

Mixed Erlang distributions are known to yield analytic solutions to many risk management problems of interest. This is primarily due to the tractable features of this distributional class. Among others, the class of mixed Erlang distributions is closed under various operations such as convolutions and Esscher transformations (e.g., Willmot and Woo 2007 and Willmot and Lin 2011). As such, risk aggregation and ruin problems can more easily be tackled under mixed Erlang assumptions (e.g., Cheung and Woo 2016, Cossette et al. 2012, and Landriault and Willmot 2009). Also, Tijms (1994) showed that the class of mixed Erlang distributions is dense in the set of all continuous and positive distributions. Therefore, we consider a moment-based approximation method which capitalizes on the aforementioned properties of the mixed Erlang distribution. More precisely, we propose to approximate a distribution with known moments by a moment-matching mixed Erlang distribution. Moment-based approximations have been extensively developed in various research areas, including performance evaluation, queueing theory, and risk theory, to name a few.

Osogami and Harchol-Balter (2006) identify the following four criteria to evaluate moment-matching algorithms: (1) *the number of moments matched;* (2) *the computational efficiency of the algorithm;* (3) *the generality of the solution;* and (4) *the minimality of the number of parameters (phases).* It also seems desirable for the *approximation to be in itself a distribution.* This is not mentioned in Osogami and Harchol-Balter (2006) for the obvious reason that they consider phase-type distributions as their moment-based approximation class. There exists an extensive literature on the approximation of distributions by a specific subset of phase-type distributions using moment-based techniques. For instance, Whitt (1982) proposed a mixture of two exponential distributions or a generalized Erlang distribution as a moment-based approximation when either the coefficient of variation (CV) is greater than or less than 1, respectively. Also, both Altiok (1985) and Vanden Bosch et al. (2000) proposed an alternative to the moment-based approximation of Whitt (1982) when CV > 1 using a Coxian distribution. Alternatively, Johnson and Taaffe (1989) considered a mixture of Erlangs with a common shape (order) parameter as their moment-based approximation.

Most predominantly, there exists a substantial body of literature on the three-moment approximation within the phase-type class of distributions (e.g., Telek and Heindl 2002, Bobbio et al. 2005, and references therein). Matching the first three moments is often viewed as effective to provide a reasonable approximation to the underlying system (e.g., Osogami and Harchol-Balter 2006 and references therein). However, as illustrated in this paper and many others, three moments does not always suffice, triggering the development of more flexible moment-based approximations. Among others, we mention the work of Johnson and Taaffe (1989) on mixed Erlang distributions of common order. Also, Dufresne (2007) proposes two approximation techniques based on Jacobi polynomial expansions and the logbeta distribution to fit combinations of exponential distributions. This paper is complementary to the aforementioned ones by considering the family of finite mixture of Erlangs with common rate parameter to approximate a distribution on $\mathbb{R}^+$, as theoretically justified in the continuous case by Tijms (1994, Theorem 3.9.1). The reader is also referred to Lee and Lin (2010) where fitting of the same class of distributions is considered using the EM algorithm (which relies on the knowledge of the approximated distribution rather than only its moments).

It is worth pointing out that other non-phase type approximation methods have been widely used in actuarial science. A good survey paper on this topic is Chaubey et al. (1998). One of these approximation classes are refinements to the normal approximation such as the normal power and the Cornish Fisher approximations (e.g., Ramsay 1991, Daykin et al. 1994, and Lee and Lin 1992). These approximations are based on the first few moments. However, the resulting approximation is often not a proper distribution. Other moment-based distributional approximations are the translated gamma distribution (e.g., Seal 1977), translated inverse Gaussian distribution (e.g.,

Chaubey et al. 1998) and the generalized Pareto distribution (e.g., Venter 1983). It should be noted that all these approximation methods are designed to fit a specific number of moments, and thus lack the flexibility to match an arbitrary number of moments.

The rest of the paper is constructed as follows. In Section 2, a brief review on admissible moments, mixed Erlang distributions and the approximation method of Johnson and Taaffe (1989) is provided. Section 3 is devoted to our class of finite mixture of Erlangs with common rate parameter. Theoretical and practical considerations related to the approximation method are drawn. Various examples are considered to examine the quality of the resulting approximation. In Section 4, we consider applications of our moment-based approximations of Section 3 when the underlying distribution is of mixed Erlang form with known rate parameter. A parallel is drawn with a discrete moment-matching problem and certain stochastic orderings, notably the *s*-convex stochastic order (e.g., Denuit et al. 1998). An application of Cossette et al. (2002) will be examined in more detail.

2. Background

2.1. Admissible moments

Karlin and Studden (1966) provide the necessary and sufficient conditions for a set of (raw) moments $\mu_m = (\mu_1, \ldots, \mu_m)$ to be from a probability distribution defined on $\mathbb{R}^+$. To state this result, define the matrices $\mathbf{P}_k$ and $\mathbf{Q}_k$ ($k \geq 1$) as

$$\mathbf{P}_k = \begin{pmatrix} 1 & \mu_1 & \cdots & \mu_k \\ \mu_1 & \mu_2 & \cdots & \mu_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_k & \mu_{k+1} & \cdots & \mu_{2k} \end{pmatrix};$$

$$\mathbf{Q}_k = \begin{pmatrix} \mu_1 & \mu_2 & \cdots & \mu_{k+1} \\ \mu_2 & \mu_3 & \cdots & \mu_{k+2} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{k+1} & \mu_{k+2} & \cdots & \mu_{2k+1} \end{pmatrix}.$$

As stated in Courtois and Denuit (2007), there exists a non-negative random variable (rv) with distribution function (df) F and first m moments μ_m if and only if the following two conditions are satisfied:

- $\det \mathbf{P}_k > 0$, for $k = 1, \ldots, \lfloor (m-1)/2 \rfloor$;
- $\det \mathbf{Q}_k > 0$, for $k = 1, \ldots, \lfloor m/2 \rfloor$;

where $\lfloor x \rfloor$ holds for the integer part of x. In what follows, we silently assume the moment set μ_m is from a probability distribution on $\mathbb{R}^+$.

2.2. Mixed Erlang distribution

We now review some known properties of mixed Erlang distributions with common rate parameter. A more elaborate review of this class of distributions can be found in Willmot and Woo (2007), Lee and Lin (2010), and Willmot and Lin (2011).

Let W be a mixed Erlang rv with common rate parameter $\beta > 0$ and df

$$F_W(x) = \sum_{k \in A_l} \zeta_k H(x; k, \beta), \tag{1}$$

where $A_l = \{1, 2, \ldots, l\}$, and $\{\zeta_k\}_{k=1}^l$ is the probability mass function (pmf) of a discrete rv K with support A_l for a given $l \in \{1, 2, \ldots\} \cup \{\infty\}$. The Erlang df H is defined as

$$H(x; k, \beta) \equiv 1 - \overline{H}(x; k, \beta) = 1 - e^{-\beta x} \sum_{i=0}^{k-1} \frac{(\beta x)^i}{i!},$$

$$x \geq 0, \tag{2}$$

where the parameters k and β of the Erlang df are known as the *shape* and *rate* parameters, respectively. An alternative and useful representation of the mixed Erlang rv W is $W = \sum_{k=1}^K C_k$ where $\{C_k\}_{k \geq 1}$ are iid exponential rv's with mean $1/\beta$, independent of K, i.e., the rv W follows a compound distribution.

Remark 1. *As in, e.g., Willmot and Woo (2007), we consider the class of mixed Erlang dfs* (1) *rather than the more general class of combinations of Erlangs where some ζ_k's are possibly negative. For the latter class, additional constraints on $\{\zeta_k\}_{k=1}^l$ exist to ensure that the right-hand side of* (1) *is a non-decreasing function in x. This presents additional challenges in*

the subsequent moment-matching application, challenges which do not arise in the mixed Erlang case.

It is well known that the j-th moment of W is given by $E[W^j] = \beta^{-j}\sum_{k=1}^{\infty}\zeta_k\{\prod_{i=0}^{j-1}(k+i)\}$. Of particular importance in actuarial science and quantitative risk management (see, e.g., McNeil et al. 2005 and references therein) are the VaR and TVaR risk measures. For the mixed Erlang rv W, there is in general no closed form expression for $\mathrm{VaR}_\kappa(W) = \inf\{x \in \mathbb{R} : F_W(x) \geq \kappa\}$ where $0 \leq \kappa < 1$, but its value can be obtained using a routine numerical procedure. As for its TVaR, Lee and Lin (2010) showed that

$$\mathrm{TVaR}_\kappa(W) \equiv \frac{1}{1-\kappa}\int_\kappa^1 \mathrm{VaR}_u(W)\,du$$

$$= \frac{1}{1-\kappa}\sum_{k=1}^{\infty}\zeta_k\frac{k}{\beta}\overline{H}(\mathrm{VaR}_\kappa(W); k+1, \beta). \quad (3)$$

Another quantity of interest is the stop-loss premium defined as $\pi_W(b) = E[(W-b)_+]$ with $(x)_+ = \max\{x, 0\}$. For the mixed Erlang df (1), we have

$$\pi_W(b) = \frac{1}{1-\kappa}\sum_{k=1}^{\infty}\zeta_k\left(\frac{k}{\beta}\overline{H}(b; k+1, \beta) - b\overline{H}(b; k, \beta)\right),$$

$$b \geq 0$$

(see also Willmot and Woo (2007, Eq. 3.6) for the higher-order stop-loss moments). Tijms (1994) showed that this class of distributions can approximate any continuous positive distribution with an arbitrary level of accuracy. For completeness, the theoretical foundation of this result is given next.

Theorem 2. *(Tijms 1994, Theorem 3.9.1). Let F be the df of a positive rv. For any given h > 0, define*

$$F_h(x) = \sum_{k=1}^{\infty}(F(kh) - F((k-1)h))H\left(x; k, \frac{1}{h}\right). \quad (4)$$

Then, $\lim_{h\to 0} F_h(x) = F(x)$ for any continuity point x of F.

Note that F_h in (4) is a mixed Erlang df of the form (1) with $\zeta_k = (F(kh) - F((k-1)h))$ $(k = 1, 2, \ldots)$ and rate parameter $\beta = 1/h$.

Several approximation methods motivated by Tijms' theorem were proposed over the years (see Section 1

for more details). In general, these moment-based approximations propose to work with a specific subclass of all finite and infinite mixed Erlang distributions. Among them, we recall the method of Johnson and Taaffe (1989), who will be used later for comparative purposes.

2.3. Method of Johnson and Taaffe (1989)

Johnson and Taaffe (1989) investigated the use of mixtures of Erlang distributions with common shape parameter for moment-matching purposes. More precisely, mixtures of n (or fewer) Erlangs with common shape parameters are used to match the first $(2n - 1)$ moments (whenever the set of moments is within the feasible set). For the three-moment matching problem, Johnson and Taaffe (1989) generalized the approximation of Whitt (1982) and Altiok (1985) by enlarging the set of feasible moments μ_3 when $CV > 1$. Their method is also valid for some combinations of μ_3 when $CV < 1$.

Their three-moment approximation is a mixture of two Erlangs with common shape parameter r (see Theorem 3 of Johnson and Taaffe 1989), i.e.,

$$F(x) = pH(x; r, \beta_1) + (1-p)H(x; r, \beta_2), \quad (5)$$

where $p = \left(\dfrac{\mu_1}{r} - \beta_2^{-1}\right)\bigg/\left(\beta_1^{-1} - \beta_2^{-1}\right)$, and $\{1/\beta_i\}_{i=1}^{2}$ are the solutions of $As^2 + Bs + C = 0$ with

$$A = r(r+2)\mu_1\left(\mu_2 - \frac{r+1}{r}\mu_1^2\right)$$

$$B = -\left(\begin{array}{c} r\left(\mu_1\mu_3 - \dfrac{r+1}{r+2}\mu_2^2\right) \\[2mm] + \dfrac{r(r+2)}{r+1}\left(\mu_2 - \dfrac{r+1}{r}\mu_1^2\right)^2 \\[2mm] + (n+2)\mu_1^2\left(\mu_2 - \dfrac{r+1}{r}\mu_1^2\right) \end{array}\right)$$

$$C = \mu_1\left(\mu_1\mu_3 - \frac{r+1}{r+2}\mu_2^2\right).$$

The choice of the shape parameter r is discussed in Johnson and Taaffe (1989, Proposition 4).

3. Moment-based approximation with mixed Erlang distribution

In this section, we propose to use a different subclass of mixed Erlang distributions to examine moment-based approximation techniques.

3.1. Description of the approach

For a given $l \in \mathbb{N}^+$, let $\mathcal{ME}(\boldsymbol{\mu}_m, A_l)$ be the set of all finite mixture of Erlangs with df (1) and first m moments $\boldsymbol{\mu}_m$. From Section 2.2, this consists in the identification of all solutions to the problem

$$\sum_{k=1}^{l} \zeta_k \frac{\prod_{i=0}^{j-1}(k+i)}{\beta^j} = \mu_j, \qquad j = 1, \ldots, m, \quad (6)$$

under the constraints that $\beta > 0$ and $\{\zeta_k\}_{k=1}^{l}$ is a probability measure on A_l.

Remark 3. *For a rv W with df (1) and first m moments $\boldsymbol{\mu}_m$, we indifferently write $W \in \mathcal{ME}(\boldsymbol{\mu}_m, A_l)$ or $F_W \in \mathcal{ME}(\boldsymbol{\mu}_m, A_l)$. This will also apply to the other distributional classes.*

Also, let $\mathcal{ME}^{res}(\boldsymbol{\mu}_m, A_l)$ be the (restricted) subset of $\mathcal{ME}(\boldsymbol{\mu}_m, A_l)$ with at most m non-zero mixing probabilities $\{\zeta_k\}_{k=1}^{l}$. Given that $\mathcal{ME}^{res}(\boldsymbol{\mu}_m, A_l)$ has a finite number of solutions, we propose to use $\mathcal{ME}^{res}(\boldsymbol{\mu}_m, A_l)$ as our approximation class. It is clear that $\mathcal{ME}^{res}(\boldsymbol{\mu}_m, A_l) \subseteq \mathcal{ME}^{res}(\boldsymbol{\mu}_m, A_{l'})$ for $l \leq l'$.

Note that for a continuous positive distribution with moments $\boldsymbol{\mu}_m$, we know from Theorem 2 that there exists a l large enough such that $\mathcal{ME}(\boldsymbol{\mu}_m, A_l)$ is not empty. Even though no formal conclusion can be reached for the restricted class $\mathcal{ME}^{res}(\boldsymbol{\mu}_m, A_l)$, all our numerical studies have shown that this set has a large number of distributions (see, for instance, the examples of subsections 3.2.1 and 3.2.2) for a given m when l is chosen large enough.

Distributions in the $\mathcal{ME}^{res}(\boldsymbol{\mu}_m, A_l)$ class are identified as follows: for a given set $\{i_k\}_{k=1}^{m} \subset A_l$ with $1 \leq i_1 < i_2 < \cdots < i_m \leq l$, (6) can be rewritten in matrix form as

$$\mathbf{G}_m \boldsymbol{\zeta}_m = \mathbf{M}_\beta,$$

where $\boldsymbol{\zeta}_m = (\zeta_{i_1}, \zeta_{i_2}, \ldots, \zeta_{i_m})^T$, $\mathbf{M}_\beta = (\beta\mu_1, \beta^2\mu_2, \ldots, \beta^m\mu_m)^T$, and

$$\mathbf{G}_m = \begin{pmatrix} i_1 & i_2 & \cdots & i_m \\ i_1(i_1+1) & i_2(i_2+1) & \cdots & i_m(i_m+1) \\ \vdots & \vdots & \ddots & \vdots \\ \prod_{i=0}^{m-1}(i_1+i) & \prod_{i=0}^{m-1}(i_2+i) & \cdots & \prod_{i=0}^{m-1}(i_m+i) \end{pmatrix}.$$

It follows that $\boldsymbol{\zeta}_m = \mathbf{G}_m^{-1}\mathbf{M}_\beta$ under the constraint that $\boldsymbol{\zeta}_m \mathbf{e} = 1$, where $\mathbf{e}$ is the vector 1's. Note that $\boldsymbol{\zeta}_m^T \mathbf{e} = (\mathbf{G}_m^{-1}\mathbf{M}_\beta)^T \mathbf{e}$ is a polynomial of degree (at most) m in β. Thus, we only consider the **real and positive** solutions (in β) of $(\mathbf{G}_m^{-1}\mathbf{M}_\beta)^T \mathbf{e} = 1$, and complete their mixed Erlang representation with the identification of the mixing weights $\boldsymbol{\zeta}_m = \mathbf{G}_m^{-1}\mathbf{M}_\beta$. The procedure is systematically repeated for all $\binom{l}{m}$ possible sets of m distinct elements in A_l.

Remark 4. *Given that the above procedure is repeated $\binom{l}{m}$ times, the computational efficiency of the proposed methodology is mostly driven by this number, and hence the parameters m and l should be chosen accordingly. For a given number of moments, m, we observe that: (a) larger values of l result in a more time-consuming numerical procedure; (b) however, l should be chosen large enough for the approximation class $\mathcal{ME}^{res}(\boldsymbol{\mu}_m, A_l)$ to have a reasonable number of members (to legitimately produce a "good" approximation). From our numerical studies, we observe that the selection of l (for a given m) can be problem-specific, and thus this tradeoff in the choice of l should be handled with care. However, as a rule of thumb, when m is relatively small (i.e., $m \leq 6$) which is traditionally in moment-matching exercises, a value of l between 50 and 100 leads to reasonable mixed Erlang approximations. We refer the reader to the numerical illustrations and subsequent remarks in Section 3.2 for a more detailed discussion on this topic.*

Among all $\mathcal{ME}^{res}(\boldsymbol{\mu}_m, A_l)$ distributions, we propose to use as a criterion of the quality of these approximations the Kolmogorov-Smirnov (KS) distance defined for two rv's S and W (with respective dfs F_S and F_W)

as $d_{KS}(S, W) = \sup_{x \geq 0} |F_S(x) - F_W(x)|$. The KS distance is commonly used in the context of continuous distributions (e.g., Denuit et al. 2005). Therefore, the chosen mixed Erlang approximation within $\mathcal{ME}^{res}(\boldsymbol{\mu}_m, A_l)$ is the one minimizing the KS distance with the true df F_S. We denote by $F_{W_{m,l}}$ this approximation, i.e.,

$$d_{KS}(S, W_{m,l}) = \inf_{F_W \in \mathcal{ME}^{res}(\boldsymbol{\mu}_m, A_l)} \sup_{x \geq 0} |F_S(x) - F_W(x)|,$$

where $W_{m,l}$ is a rv with df $F_{W_{m,l}}$. This requires the calculation of the KS distance for each mixed Erlang distribution in $\mathcal{ME}^{res}(\boldsymbol{\mu}_m, A_l)$ to identify its minimizer $W_{m,l}$. In general, an explicit expression for this KS distance does not exist, and hence we propose to numerically find this value by evaluating the distance between the two dfs over all multiples (up to a given high value) of a small discretization span.

Note that other distances such as the stop-loss distance (e.g., Gerber 1979) could have been used to select our approximation distribution. Alternatively, one could have relied on another criterion to identify this approximation distribution (for instance, select the distribution in $\mathcal{ME}^{res}(\boldsymbol{\mu}_m, A_l)$ with the closest subsequent moment to the true distribution).

3.2. Numerical examples

We consider a few simple examples to illustrate the quality of the approximation. For comparative purposes, other approximation methods will also be discussed. Some concluding remarks on the mixed Erlang approximation method are later made based on the numerical experiment conducted next.

3.2.1. Lognormal distribution: Dufresne (2007, Example 5.4)

Let $S = \exp(Z)$ where Z is a normal rv with mean 0 and variance 0.25. The first 5 moments of S are $\boldsymbol{\mu}_5 = (1.1331, 1.6487, 3.0802, 7.3891, 22.7599)$. We consider the class of mixed Erlang distributions $\mathcal{ME}^{res}(\boldsymbol{\mu}_m, A_{70})$ ($m = 3, 4, 5$) which have a total of 13198, 89294 and 290422 distributions, respectively. The resulting mixed Erlang approximations $F_{W_{m,70}}$ ($m = 3, 4, 5$) are

$$F_{W_{3,70}}(x) = 0.8209H(x; 6, 6.3219)$$
$$+ 0.1727H(x; 12, 6.3219)$$
$$+ 0.0064H(x; 26, 6.3219),$$

$$F_{W_{4,70}}(x) = 0.6350H(x; 7, 8.3334)$$
$$+ 0.2950H(x; 12, 8.3334)$$
$$+ 0.0672H(x; 20, 8.3334)$$
$$+ 0.0029H(x; 40, 8.3334),$$

and

$$F_{W_{5,70}}(x) = 0.6273H(x; 7, 8.3608)$$
$$+ 0.3063H(x; 12, 8.3608)$$
$$+ 0.0609H(x; 20, 8.3608)$$
$$+ 0.0055H(x; 34, 8.3608)$$
$$+ 0.0001H(x; 69, 8.3608),$$

for $x \geq 0$ with respective KS distances of $d_{KS}(S, W_{3,70}) = 0.0040$, $d_{KS}(S, W_{4,70}) = 0.0018$ and $d_{KS}(S, W_{5,70}) = 0.0011$. Note that the quality of the mixed Erlang approximation (as measured by the KS distance) increases with the number of moments matched. For comparative purposes, the three-moment approximation (5) of Johnson and Taaffe (1989) is given by

$$F_{W_{JT}}(x) = 0.0087H(x; 4, 1.2804)$$
$$+ 0.9913H(x; 4, 3.5855),$$

In Figure 1, we compare the density function of $W_{m,70}$ ($m = 3, 4, 5$) W_{JT} and S.

All three mixed Erlang approximations provide an overall good fit to the exact distribution. To further examine the tail fit, specific values of VaR and TVaR for the exact and approximated distributions are provided in Tables 1 and 2, respectively.

We observe that the VaR and TVaR values of the mixed Erlang approximations compare very well to their lognormal counterparts, especially for the 5-moment approximation. This is particularly true given that the lognormal distribution is known to have a heavier tail than the mixed Erlang distribution.

Figure 1. Density function: Lognormal vs. Approximations
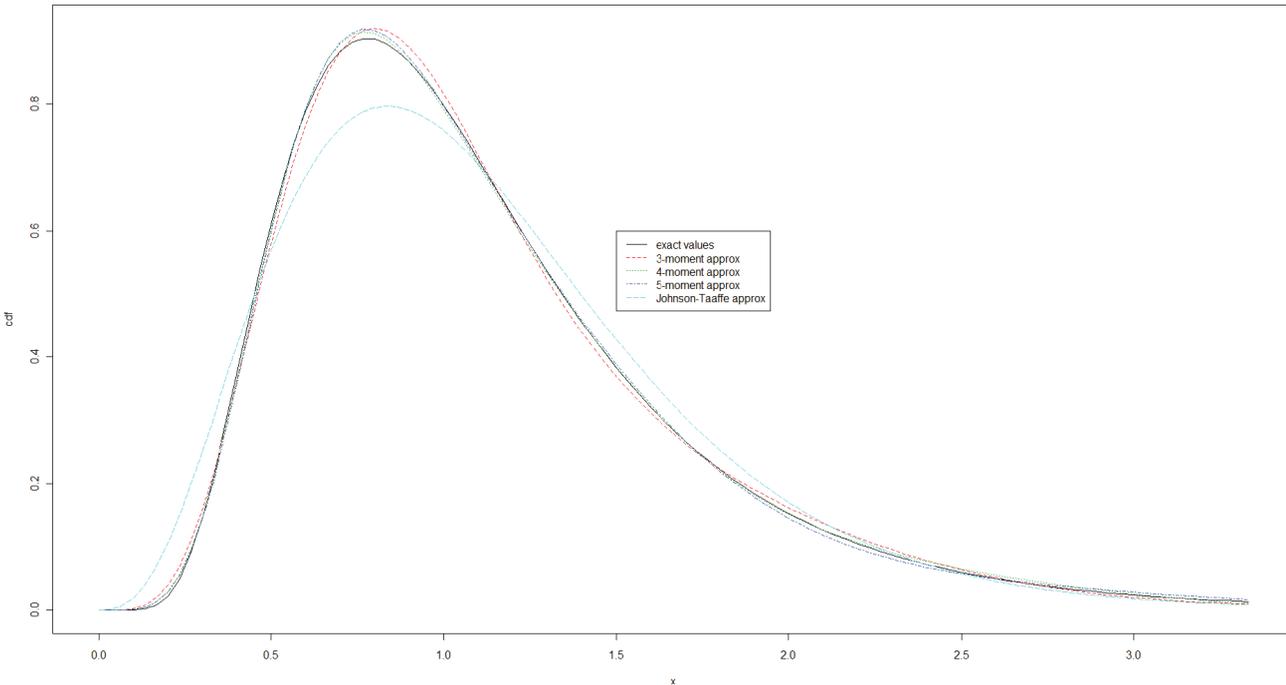
Note that the improvement is indeed not monotone with the number of moments matched, as increasing this number does not necessarily lead to a higher quality approximation in moment-matching techniques.

3.2.2. Mixture of two gamma distributions: Lee and Lin (2010, Section 5, Example 1)

Let S be a mixture of two gamma distributions with density

$$f_S(s) = 0.2 \frac{(3.2s)^{2.6} e^{-3.2s}}{s\Gamma(2.6)}$$

$$+ 0.8 \frac{(1.2s)^{6.3} e^{-1.2s}}{s\Gamma(6.3)}, \qquad s \geq 0,$$

and first 6 moments $\mu_6 = (4.3623, 25.7308, 176.9624, 1369.8272, 11754.2149, 110674.4154)$. We consider the class of mixed Erlang distributions $\mathcal{ME}^{res}(\mu_m, A_{70})$

Table 1. Values of VaR$_\kappa$ for W_{JT}, $W_{m,70}$ ($m = 3, 4, 5$) and S

κ	VaR$_\kappa$ (W_{JT})	VaR$_\kappa$ ($W_{3,70}$)	VaR$_\kappa$ ($W_{4,70}$)	VaR$_\kappa$ ($W_{5,70}$)	VaR$_\kappa$ (S)
0.9	1.8906	1.9129	1.9056	1.8936	1.8980
0.95	2.2119	2.2692	2.2918	2.2791	2.2760
0.99	2.9953	3.1223	3.0991	3.1812	3.2001
0.995	3.4239	3.6892	3.4811	3.6572	3.6252
0.999	5.0672	4.9237	5.0623	4.7241	4.6885

Table 2. Values of TVaR$_\kappa$ for W_{JT}, $W_{m,70}$ ($m = 3, 4, 5$) and S

κ	TVaR$_\kappa$ (W_{JT})	TVaR$_\kappa$ ($W_{3,70}$)	TVaR$_\kappa$ ($W_{4,70}$)	TVaR$_\kappa$ ($W_{5,70}$)	TVaR$_\kappa$ (S)
0.9	2.3931	2.4540	2.4601	2.4616	2.4616
0.95	2.7528	2.8350	2.8431	2.8600	2.8586
0.99	3.7939	3.9007	3.8125	3.8579	3.8413
0.995	4.4115	4.4455	4.3654	4.3323	4.2957
0.999	6.2260	5.4245	5.6195	5.3151	5.4341

Table 3. Values of VaR$_\kappa$ for $W_{m,70}$ (m = 3, 4, 5, 6), W_{EM}, and S

κ	VaR$_\kappa$ (W_{EM})	VaR$_\kappa$ ($W_{3,70}$)	VaR$_\kappa$ ($W_{4,70}$)	VaR$_\kappa$ ($W_{5,70}$)	VaR$_\kappa$ ($W_{6,70}$)	VaR$_\kappa$ (S)
0.9	7.7069	7.8183	7.6726	7.6943	7.6969	7.6859
0.95	8.7494	8.8676	8.7595	8.7172	8.7825	8.7666
0.99	10.7708	10.8451	11.0385	11.0462	10.9514	11.0023
0.995	11.5349	11.5835	11.9366	12.0611	11.8566	11.8925
0.999	13.1604	13.1473	13.8488	13.9701	13.9651	13.8551

for $m = 3, 4, 5, 6$ which are composed of 16000, 83797, 494532 and 1928919 distributions, respectively. The resulting mixed Erlang approximations $F_{W_{m,70}}$ ($m = 3, 4, 5, 6$) are

$$F_{W_{3,70}}(x) = 0.2140H(x; 2, 2.0835)$$

$$+ 0.5215H(x; 9, 2.0835)$$

$$+ 0.2645H(x; 15, 2.0835),$$

$$F_{W_{4,70}}(x) = 0.2266H(x; 2, 2.0469)$$

$$+ 0.4440H(x; 9, 2.0469)$$

$$+ 0.2963H(x; 13, 2.0469)$$

$$+ 0.0330H(x; 19, 2.0469),$$

$$F_{W_{5,70}}(x) = 0.2023H(x; 3, 3.7271)$$

$$+ 0.2091H(x; 12, 3.7271)$$

$$+ 0.3936H(x; 19, 3.7271)$$

$$+ 0.1805H(x; 28, 3.7271)$$

$$+ 0.0145H(x; 42, 3.7271),$$

and

$$F_{W_{6,70}}(x) = 0.0768H(x; 2, 3.0731)$$

$$+ 0.1325H(x; 3, 3.0731)$$

$$+ 0.2739H(x; 11, 3.0731)$$

$$+ 0.3928H(x; 17, 3.0731)$$

$$+ 0.1188H(x; 25, 3.0731)$$

$$+ 0.0052H(x; 37, 3.0731),$$

with respective KS distances of $d_{KS}(S, W_{3,70}) = 0.0148$, $d_{KS}(S, W_{4,70}) = 0.0050$, $d_{KS}(S, W_{5,70}) = 0.0035$ and $d_{KS}(S, W_{6,70}) = 0.0024$. Lee and Lin (2010) used the EM algorithm to fit a mixed Erlang distribution to the same distribution, which resulted in the following model:

$$F_{W_{EM}}(x) = 0.2282H(x; 2, 1.9603)$$

$$+ 0.5430H(x; 9, 1.9603)$$

$$+ 0.2288H(x; 14, 1.9603),$$

with KS distance $d_{KS}(S, W_{EM}) = 0.0094$. Note that the KS distance for the 3-moment approximation is greater than the KS distance obtained using the EM estimation. We recall that the EM algorithm finds maximum likelihood estimates using the approximated distribution as an input, while our method is based on only partial information on the approximated distribution (e.g., its first m moments). We observe that the fit improves and the KS distance decreases when more moments are included in the approximation. For illustrative purposes, we also provide in Tables 3 and 4

Table 4. Values of TVaR$_\kappa$ for $W_{m,70}$ (m = 3, 4, 5, 6), W_{EM} and S

κ	TVaR$_\kappa$ (W_{EM})	TVaR$_\kappa$ ($W_{3,70}$)	TVaR$_\kappa$ ($W_{4,70}$)	TVaR$_\kappa$ ($W_{5,70}$)	TVaR$_\kappa$ ($W_{6,70}$)	TVaR$_\kappa$ (S)
0.9	9.0866	9.1909	9.1596	9.1404	9.1634	9.1598
0.95	9.9929	10.0845	10.1583	10.1234	10.1378	10.1469
0.99	11.8253	11.8620	12.2766	12.3665	12.2440	12.2528
0.995	12.5377	12.5481	13.1132	13.2302	13.1371	13.1065
0.999	14.0774	14.0266	14.9062	14.8770	15.0978	15.0069

some values of VaR and TVaR for $W_{m,70}$ ($m = 3, 4, 5, 6$), W_{EM} and S.

3.2.3. Gompertz distribution

The Gompertz distribution with df $F_S(x) = 1 - \exp\{-B(c^x - 1)/\ln c\}$ ($x > 0$) has been extensively applied in various life contingency contexts (e.g., Bowers et al. 1997). Lenart (2014) provides an expression for the j-th moment, namely

$$\mu_j = \frac{j!}{(\ln c)^j} e^{\frac{B}{\ln c}} E_1^{j-1}\left(\frac{B}{\ln c}\right), \qquad (7)$$

where $E_s^j(z) = \int_1^\infty \frac{(\ln x)^j}{j!} x^{-s} e^{-zx} dx$ is the generalized integro-exponential function (see Milgram 1985). Here, we consider an example of Melnikov and Romaniuk (2006) on the 1959–1999 USA mortality data of the human mortality database where the parameters B and c were estimated to $B = 6.148 \times 10^{-5}$ and $c = 1.09159$. Using (7), the first five moments are $\mu_5 = (76.3437, 6037.202, 489676.3, 40524308, 3410245408)$. We consider the class of mixed Erlang distributions $\mathcal{ME}^{res}(\mu_m, A_{90})$. The resulting mixed Erlang approximations $F_{W_{3,90}}$ and $F_{W_{4,90}}$ are

$$F_{W_{3,90}}(x) = 0.0154 H(x; 22, 1.0928)$$
$$+ 0.2210 H(x; 65, 1.0928)$$
$$+ 0.7637 H(x; 90, 1.0928)$$

and

$$F_{W_{4,90}}(x) = 0.0347 H(x; 35, 1.0972)$$
$$+ 0.0834 H(x; 66, 1.0972)$$
$$+ 0.1009 H(x; 67, 1.0972)$$
$$+ 0.7810 H(x; 90, 1.0972),$$

with KS distances $d_{KS}(S, W_{3,90}) = 0.0188$, and $d_{KS}(S, W_{4,90}) = 0.0239$. In Figure 2, we compare the fit of the 3- and 4-moment approximations to the exact distribution by plotting their densities (left) and dfs (right). Overall, we observe that the fit is quite reasonable.

Note that the KS distance increases from the 3-moment to the 4-moment approximation. We notice that both $F_{W_{3,90}}$ and $F_{W_{4,90}}$ use the Erlang-90 df, where 90 is the largest element of A_{90}. As such, a mixed Erlang approximation with a smaller KS distance can likely be found in both cases by choosing a larger support A_l (i.e., $l > 90$).

3.2.4. Some remarks

To provide insight on the quality of the moment-based mixed Erlang approximation proposed in this section, we briefly revisit the results of the above three examples. In the first two (lognormal and mixture of two gammas), the mixed Erlang approximation is easy to implement and provides a very satisfactory fit

Figure 2. Density (left) and df (right): 3- and 4-moment approximations vs. Gompertz distribution
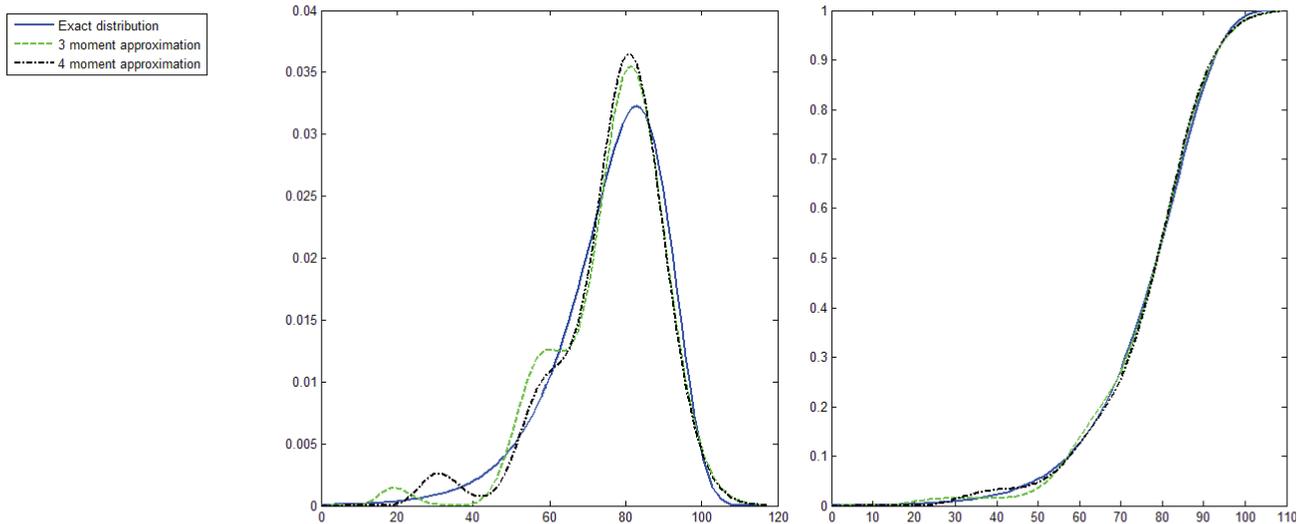
to the true distribution. As none of the resulting mixed Erlang approximations uses the Erlang-70 df (as $l = 70$ in the first two examples), it is unlikely that a better approximation (from the viewpoint of KS distance) can be found by increasing the value of l. Given that the best KS-fit is found from a mixture of Erlang distributions with relatively small shape parameter (which corresponds to the parameter k in the Erlang df (2), the proposed method seems particularly well suited for these two cases.

As for any approximation method, limitations can also be found, as evidenced by the Gompertz example. Indeed, for this example, both the 3-moment and 4-moment mixed Erlang approximations use the Erlang-90 df (recall $l = 90$ for this example). As mentioned earlier, one can likely reduce the KS distance (if so desired) of the resulting mixed Erlang approximation by increasing the value of l (in light of the comments in Remark 4). This implies that the KS-optimal mixed Erlang approximation would likely involve Erlang distributions with large shape parameters (which have smaller variances for a given rate parameter β).

In general, distributions with negative skewness and sharp density peak(s) may require the use of Erlang distributions with large shape parameters to provide a good approximation. Computational time of the proposed mixed Erlang methodology may become a non-negligible issue in these cases (especially as the number of moments matched increases). However, a slight adjustment to the proposed methodology may be considered to address this time-consuming issue. Indeed, one may replace the set $A_l = \{1, 2, \ldots, l\}$ in (1) by a set of the form $\{a, 2a, \ldots, ja\}$ for positive integers a and j (note that the two sets coincide when $a = 1$ and $j = l$). To illustrate this, we have re-considered the Gompertz example by replacing the set A_{90} by the set $\{5, 10, 15, \ldots, 200\}$. The resulting mixed Erlang approximation, denoted by the rv $W_{m,5:200}$ when m moments are matched ($m = 3,4,5$), are given by

$$F_{W_{3,5:200}}(x) = 0.0447H(x; 45, 1.3185)$$
$$+ 0.2573H(x; 85, 1.3185)$$
$$+ 0.6980H(x; 110, 1.3185),$$

$$F_{W_{4,5:200}}(x) = 0.0168H(x; 40, 1.6422)$$
$$+ 0.0971H(x; 85, 1.6422)$$
$$+ 0.3044H(x; 115, 1.6422)$$
$$+ 0.5818H(x; 140, 1.6422),$$

and

$$F_{W_{5,5:200}}(x) = 0.0038H(x; 20, 1.9095)$$
$$+ 0.0393H(x; 75, 1.9095)$$
$$+ 0.1262H(x; 110, 1.9095)$$
$$+ 0.3275H(x; 140, 1.9095)$$
$$+ 0.5032H(x; 165, 1.9095),$$

with KS distances $d_{KS}(S, W_{3,5:200}) = 0.0171$, $d_{KS}(S, W_{4,5:200}) = 0.0086$, and $d_{KS}(S, W_{5,5:200}) = 0.0072$. Note that the KS distance in the 4-moment approximation is considerably lower than for $d_{KS}(S, W_{4,90}) = 0.0239$. In Figure 3, we compare the fit of the 5-moment approximation $W_{5,5:200}$ to the Gompertz distribution by plotting their densities (left) and dfs (right). We can see that the fit is quite acceptable and of a better quality than the two approximations displayed in Figure 3.

4. Moment-based mixed Erlang approximation with known β

4.1. Basic definitions

We consider here a slightly different context than the one of Section 3. Instead of approximating a general df F_S with known moments μ_m, we assume that the df F_S is known to be of mixed Erlang form (1) with given rate parameter $\beta > 0$, and first m moments μ_m. However, the mixing weights $\{\zeta_k\}_{k=1}^l$ are assumed unknown or difficult to obtain. Various applications in risk theory and credit risk fall into this context (e.g., Cossette et al. 2002, Lindskog and McNeil 2003, and McNeil et al. 2005; see also the application of Section 4.4).

For a given rate parameter $\beta > 0$, let $\mathcal{ME}(\mu_m, \beta)$ be the set of all mixed Erlang distributions with df (1) (as $l \to \infty$) and first m moments μ_m. Also, define $\mathcal{ME}(\mu_m, A_l, \beta)$ to be the subset of $\mathcal{ME}(\mu_m, \beta)$ with

Figure 3. Density (left) and df (right): 5-moment approximation vs Gompertz distribution
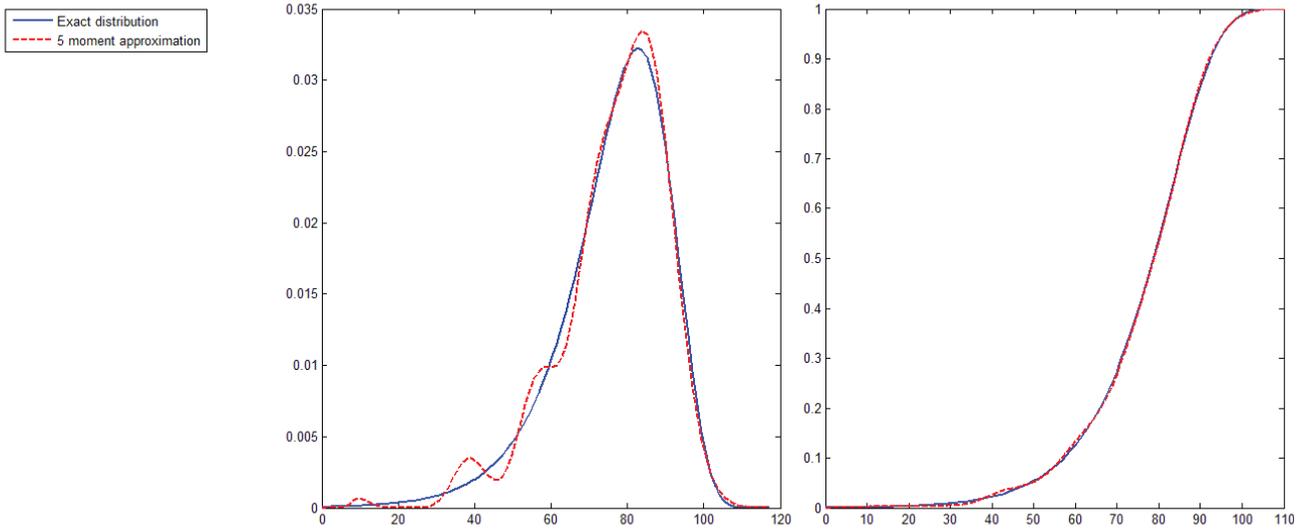
df (1) for a given $l \in \mathbb{N}^+$, and let $\mathcal{ME}^{ext}(\mu_m, A_l, \beta)$ be a further subset of $\mathcal{ME}(\mu_m, A_l, \beta)$ such that at most $(m + 1)$ of the mixing weights $\{\zeta_k\}_{k=1}^l$ are non-zero. Note that a distribution in $\mathcal{ME}(\mu_m, A_l, \beta)$ can be expressed as a convex combination of distributions in $\mathcal{ME}^{ext}(\mu_m, A_l, \beta)$ (see, e.g., De Vylder 1996).

For a given function ϕ, we consider two approaches to derive bounds and approximations for $E[\phi(S)]$ when $S \in \mathcal{ME}(\mu_m, A_l, \beta)$ (in cases when the expectation exists). The first approach is based on discrete s-convex extremal distributions while the second is based on moment bounds on discrete expected stop-loss transforms.

Remark 5. *Naturally, the set $\mathcal{ME}(\mu_m, A_l, \beta)$ tends to $\mathcal{ME}(\mu_m, \beta)$ as $l \to \infty$. As such, when $S \in \mathcal{ME}(\mu_m, \beta)$, bounds for risk measures on $\mathcal{ME}(\mu_m, \beta)$ can be approximated by their counterparts in $\mathcal{ME}(\mu_m, A_l, \beta)$ for l reasonably large.*

For a mixed Erlang rv S with df (1), its j-th moment is known to satisfy (6). Using the identity

$$\prod_{i=0}^{j-1}(k+i) = \sum_{n=1}^{j} s(j,n)k^n,$$

where the $s(j, n)$'s are the (signed) Stirling numbers of the first kind (e.g., Abramowitz and Stegun 1972), (6) becomes

$$\beta^j \mu_j = \sum_{n=1}^{j} s(j,n)\kappa_n, \tag{8}$$

for $j = 1, \ldots, m$ where $\kappa_n = \sum_{k=1}^{l} \zeta_k k^n$. In matrix form, we have $\mathbf{M}_\beta = \mathbf{s}\kappa_m^T$, where $\mathbf{s} = \{s(j, n)\}_{j,n=1}^m$ (with $s(j, n) = 0$ for $n > j$) and $\kappa_m = (\kappa_1, \ldots, \kappa_m)$. Isolating κ_m yields

$$\kappa_m = \left(\mathbf{s}^{-1}\mathbf{M}_\beta\right)^T. \tag{9}$$

From Comtet (1974, p. 213) we know that $\mathbf{s}^{-1} \equiv \mathbf{c} = \{(-1)^{i+j} c(i, j)\}_{i,j=1}^m$ where the $c(i, j)$'s are the Stirling numbers of the second kind defined as $c(i, j) = (j!)^{-1}\sum_{k=0}^{j}(-1)^{j-k}\binom{j}{k}k^i$ (e.g., Abramowitz and Stegun 1972).

Thus, for a given $\beta > 0$, the class $\mathcal{ME}(\mu_m, A_l, \beta)$ can be found through the identification of all discrete distributions with support A_l and first m moments given by the right-hand side of (9). Equivalently, the class $\mathcal{ME}^{ext}(\mu_m, A_l, \beta)$ can be identified by restricting the discrete distributions on A_l to have at most $(m + 1)$ non-zero mass points. This argument is formalized in the next section.

4.2. Discrete s-convex extremal distributions

Let $\mathcal{D}(\alpha_m, A_l)$ be all discrete distributions with support A_l and first m moments $\alpha_m = (\alpha_1, \alpha_2, \ldots, \alpha_m)$. Also, denote by $\mathcal{D}^{ext}(\alpha_m, A_l)$ the subset of $\mathcal{D}(\alpha_m, A_l)$ with distributions having at most $(m + 1)$ non-zero mass points. For a given $\beta > 0$, each distribution

in $\mathcal{D}(\kappa_m, A_l)$ (and $\mathcal{D}^{ext}(\kappa_m, A_l)$) with κ_m as defined in (9) corresponds to a mixed Erlang distribution in $\mathcal{ME}(\mu_m, A_l, \beta)$ (and $\mathcal{ME}^{ext}(\mu_m, A_l, \beta)$). This is a one-to-one correspondence (e.g., De Vylder 1996, part 2).

Remark 6. *Because of this one-to-one correspondence between $\mathcal{D}(\kappa_m, A_l)$ and $\mathcal{ME}(\mu_m, A_l, \beta)$, conditions under which $\mathcal{ME}(\mu_m, A_l, \beta)$ is not empty can be found from its discrete counterpart $\mathcal{D}(\kappa_m, A_l)$. We refer the reader to, e.g., De Vylder (1996), Marceau (1996), or Courtois and Denuit (2009).*

This allows us to make use of the theory developed in Prékopa (1990), Denuit and Lefèvre (1997), Denuit, Lefèvre and Mesfioui (1999), and Courtois et al. (2006) to derive bounds/approximations for $E[\phi(S)]$ when $S \in \mathcal{ME}(\mu_m, A_l, \beta)$.

First, we briefly recall the definitions of *s*-convex function and *s*-convex order introduced by Denuit et al. (1998).

Definition 7. *Denuit, Lefèvre and Shaked (1998, 2000). Let $\mathcal{C}$ be a subinterval of $\mathbb{R}$ or a subset of $\mathbb{N}$, and ϕ a function on $\mathcal{C}$. For a positive integer s and $x_0 < x_1 < \cdots < x_s \in \mathcal{C}$, we recursively define the divided differences as*

$$[x_0, x_1, \ldots, x_k]\phi = \frac{[x_1, x_2, \ldots, x_k]\phi - [x_0, x_1, \ldots, x_{k-1}]\phi}{x_k - x_0}$$

$$= \sum_{i=1}^{k} \frac{\phi(x_i)}{\prod_{j=0, j \neq i}^{k}(x_i - x_j)}, \quad k = 1, 2, \ldots, s,$$

where $[x_k]\phi = \phi(x_k)$ for $k = 0, 1, \ldots, s$. The function ϕ is *s*-convex if $[x_0, x_1, \ldots, x_s]\phi \geq 0$ for all $x_0 < x_1 < \cdots < x_s \in \mathcal{C}$.

We mention that the definition of *s*-convex function, which refers to higher-convexity, should not be confused with the one for Schur-convex function, also known as S-convex function.

Definition 8. *Denuit, Lefèvre and Shaked (2000). For two rv's X and Y defined on $\mathcal{C}$, X is smaller than Y in the s-convex sense, namely $X \square_{s-cx}^{\mathcal{C}} Y$, if $E[\phi(X)] \leq E[\phi(Y)]$ for all s-convex functions ϕ (provided the expectation exists).*

We mention that the 1-convex order corresponds to the usual stochastic dominance order, and the 2-convex order is the usual convex order (see Müller and Stoyan (2002), Denuit et al. (2005), and Shaked and Shanthikumar (2007) for a review on stochastic orders). Also, as stated in Theorem 1.6.3 of Müller and Stoyan (2002), the *s*-convex order can only be used to compare rv's with the same first $(s - 1)$ moments (which explains why *s* is chosen to be $m + 1$ in what follows). Examples of *s*-convex functions are $\phi(x) = x^{s+j}$ for $j \in \mathbb{N}$ and $\phi(x) = \exp(cx)$ for $c \geq 0$. For a general treatment of the *s*-convex order, see, e.g., Denuit, Lefèvre, and Mesfioui (1999) and Denuit, Lefèvre, and Shaked (2000) and section 1.6 of Müller and Stoyan (2002).

Let $K_{(m+1)-\min}$ and $K_{(m+1)-\max}$ be the $(m + 1)$-extremum rv's on $\mathcal{D}(\kappa_m, A_l)$, i.e., those which satisfy

$$E[\phi(K_{(m+1)-\min})] \leq E[\phi(K)]$$

$$\leq E[\phi(K_{(m+1)-\max})],$$

for any $(m + 1)$-convex function ϕ and $K \in \mathcal{D}(\kappa_m, A_l)$. The general distribution forms of $K_{(m+1)-\min}$ and $K_{(m+1)-\max}$ are given in Prékopa (1990) (see also Courtois et al. (2006, Section 4)), and are repeated here:

	$m + 1$ even	$m + 1$ odd
support of $K_{(m+1)-\min}$	$\left\{j_1, j_1 + 1, \ldots, j_{\frac{m+1}{2}}, j_{\frac{m+1}{2}} + 1\right\}$	$\left\{1, j_1, j_1 + 1, \ldots, j_{\frac{m}{2}}, j_{\frac{m}{2}} + 1\right\}$
support of $K_{(m+1)-\max}$	$\left\{1, j_1, j_1 + 1, \ldots, j_{\frac{m-1}{2}}, j_{\frac{m-1}{2}} + 1, l\right\}$	$\left\{j_1, j_1 + 1, \ldots, j_{\frac{m}{2}}, j_{\frac{m}{2}} + 1, l\right\}$

$$(10)$$

where $1 < j_1 < j_1 + 1 < j_2 < \cdots < l$. From (10), it is clear that the support of $K_{(m+1)-\min(\max)}$ has at most $m + 1$ elements.

Let $W_K = \sum_{j=1}^{K} C_j$ be a mixed Erlang rv, i.e., $\{C_j\}_{j \geq 1}$ are a sequence of iid exponential rv's with mean $1/\beta$, independent of K. The following result of Denuit, Lefèvre and Utev (1999, Property 5.7) relates to the stability of the s-convex order under compounding.

Lemma 9 *If $K \,\square_{s-cx}^{A_l}\, K'$, then $W_K \,\square_{s-cx}^{\mathbb{R}+}\, W_{K'}$.*

We apply Lemma 9 to define the mixed Erlang rv's $W_{K_{(m+1)-\min}}$ and $W_{K_{(m+1)-\max}}$, which are the $(m + 1)$-extremum rv's on $\mathcal{ME}(\mathbf{\mu}_m, A_l, \beta)$. It is immediate that, for $W \in \mathcal{ME}(\mathbf{\mu}_m, A_l, \beta)$,

$$W_{K_{(m+1)-\min}} \preceq_{(m+1)-cx}^{\mathbb{R}+} W \preceq_{(m+1)-cx}^{\mathbb{R}+} W_{K_{(m+1)-\max}}. \quad (11)$$

For instance, using (11), the $(m + 1)$-convex functions $\phi(x) = x^{m+1+j}$ ($j \in \mathbb{N}$) and $\phi(x) = \exp(cx)$ ($c \geq 0$) yield

$$E\left[W_{K_{(m+1)-\min}}^{m+1+j}\right] \leq E\left[W^{m+1+j}\right] \leq E\left[W_{K_{(m+1)-\max}}^{m+1+j}\right]$$

and

$$E\left[\exp\left(cW_{K_{(m+1)-\min}}\right)\right] \leq E\left[\exp(cW)\right]$$

$$\leq E\left[\exp\left(cW_{K_{(m+1)-\max}}\right)\right],$$

respectively.

4.3. Moment bounds on discrete expected stop-loss transforms

Extrema on the $(m+1)$-convex order yield bounds for $E[\phi(W)]$ when ϕ is $(m+1)$-convex and $W \in \mathcal{ME}(\mathbf{\mu}_m, A_l, \beta)$. However, this approach is not appropriate to derive bounds on TVaR and the stop-loss premium when the number of known moments is greater than 2. It is well known that two rv's with the same mean and variance cannot be compared under the convex order.

Consequently, we use an approach inspired from Courtois and Denuit (2009) (see also Hürlimann 2002) to derive bounds on TVaR and the stop-loss premium. We consider $\mathcal{D}(\mathbf{\kappa}_m, A_l)$ and determine lower and upper bounds for $E[(K - k)_+]$ for all $k \in A_l$. From the lower

bound, we define the corresponding rv K_{m-down} on A_l via the df

$$F_{K_{m-down}}(k) = \begin{cases} 1 - \begin{pmatrix} \inf_{K \in \mathcal{D}(\mathbf{\kappa}_m, A_l)} E\left[(K - k)_+\right] \\ - \inf_{K \in \mathcal{D}(\mathbf{\kappa}_m, A_l)} E\left[(K - k - 1)_+\right] \end{pmatrix}, \\ \qquad\qquad\qquad k = 1, 2, \ldots, l - 1, \\ 1, \qquad\qquad\qquad k = l \end{cases}$$
$$(12)$$

for $k \in A_l$. Similarly, K_{m-up} is defined as in (12) by replacing 'inf' by 'sup' in the definition. Given that, for $K \in \mathcal{D}(\mathbf{\kappa}_m, A_l)$,

$$E\left[(K_{m-down} - k)_+\right] \leq E\left[(K - k)_+\right] \leq E\left[(K_{m-up} - k)_+\right],$$

for all $k \in A_l$, it implies that $K_{m-down(up)}$ is smaller (larger) than K under the increasing convex order, namely $K_{m-down} \,\square_{icx}\, K \,\square_{icx}\, K_{m-up}$ (see, e.g., Courtois and Denuit 2009). Note that K_{m-down} and K_{m-up} do not belong to $\mathcal{D}(\mathbf{\kappa}_m, A_l)$, but both have first moment κ_1. The increasing convex order is stable under compounding, and thus

$$W_{K_{m-down}} \preceq_{icx} W_K \preceq_{icx} W_{K_{m-up}}. \quad (13)$$

Then, from Denuit et al. (2005, Proposition 3.4.8), it follows that

$$\mathrm{TVaR}_{\kappa}\left(W_{K_{m-down}}\right) \leq \mathrm{TVaR}_{\kappa}(W_K)$$

$$\leq \mathrm{TVaR}_{\kappa}\left(W_{K_{m-up}}\right), \quad (14)$$

for $\kappa \in (0, 1)$. Clearly, the rv's $W_{K_{m-down}}$ and $W_{K_{m-up}}$ will most likely not belong to $\mathcal{ME}(\mathbf{\mu}_m, A_l, \beta)$.

Remark 10. *Note that, when neither of the two aforementioned approaches are applicable, we propose to derive approximate bounds for $E[\phi(S)]$ with $S \in \mathcal{ME}^{ext}(\mathbf{\mu}_m, A_l, \beta)$ by calculating $E[\phi(W)]$ for all $W \in \mathcal{ME}^{ext}(\mathbf{\mu}_m, A_l, \beta)$ and choosing*

$$E[\phi(W)]_{\min(\max)} = \inf_{W \in \mathcal{ME}^{ext}(\mathbf{\mu}_m, A_l, \beta)} (\sup)\, E[\phi(W)].$$

Obviously, $E[\phi(S)]$ does not necessarily lie between $E[\phi(W)]_{\min}$ and $E[\phi(W)]_{\max}$. However, the 'interval'

estimate $[E[\phi(W)]_{\min}, E[\phi(W)]_{\max}]$ may give an idea of the variability of all solutions on $\mathcal{ME}(\boldsymbol{\mu}_m, A_l, \beta)$.

4.4. Example: Portfolio of dependent risks

We consider a portfolio of n dependent risks as described in the common mixture model of Cossette et al. (2002). Let $S = X_1 + \cdots + X_n$ be the aggregate claim amount with $X_i = B_i I_i$. Conditional on a common mixture rv Θ with pmf $a_j = \mathbb{P}(\Theta = j)$ for $j = 1, 2, \ldots$, $\{I_i\}_{i=1}^n$ are assumed to form a sequence of independent Bernoulli rv's with $\mathbb{P}(I_i = 1|\Theta = j) = 1 - (r_i)^j$ for $r_i \in (0, 1)$. As for the B_i's, they are assumed to form a sequence of iid exponential rv's of mean 1, independent of $\{I_i\}_{i=1}^{20}$ and Θ.

In this context, it is clear that S is a two-point mixture of a degenerate rv at 0 and a mixed Erlang rv of the form (1) with $l = n$ and $\beta = 1$, i.e., its Laplace transform is given by

$$E\left[e^{-tS}\right] \equiv 1 - p + pE\left[e^{-tY}\right]$$

$$= \sum_{j=1}^{\infty} a_j \left\{ \prod_{i=1}^{n} \left((r_i)^j + \left(1 - (r_i)^j\right)\frac{1}{1+t}\right)\right\}, \quad t \geq 0,$$

where $p = 1 - \sum_{j=1}^{\infty} a_j \prod_{i=1}^n (r_i)^j$. We perform the moment-based approximation on the rv $Y = (S|S > 0)$ rather than S.

For illustrative purposes, we assume $n = 20$ and a logarithmic pmf for Θ, namely $a_j = (0.5)^j/(j \ln 2)$ for $j \geq 1$. Also, the constants r_i are set such that the (unconditional) mean of I_i is $q_i = 1 - E[(r_i)^\Theta]$ with $q_1 = \cdots = q_{10} = 0.1$ and $q_{11} = \cdots = q_{20} = 0.02$. Under the above assumptions, the first five moments of Y are found to be $\boldsymbol{\mu}_5 = (1.7999, 6.2270, 31.4785, 208.1258, 1693.7077)$.

Using the approach on discrete $(m + 1)$-convex extremal distributions, the dfs $F_{W_{K_{(m+1)-\min}}}$ and $F_{W_{K_{(m+1)-\max}}}$ for $m = 4, 5$ are:

$$F_{W_{K_{5-\min}}}(x) = 0.5365H(x; 1, 1)$$
$$+ 0.2127H(x; 2, 1)$$
$$+ 0.2232H(x; 3, 1)$$
$$+ 0.0245H(x; 6, 1)$$
$$+ 0.0030H(x; 7, 1),$$

$$F_{W_{K_{5-\max}}}(x) = 0.4860H(x; 1, 1)$$
$$+ 0.3981H(x; 2, 1)$$
$$+ 0.0621H(x; 4, 1)$$
$$+ 0.0537H(x; 5, 1)$$
$$+ 0.0000H(x; 20, 1),$$

$$F_{W_{K_{6-\min}}}(x) = 0.5126H(x; 1, 1)$$
$$+ 0.3179H(x; 2, 1)$$
$$+ 0.0531H(x; 3, 1)$$
$$+ 0.1082H(x; 4, 1)$$
$$+ 0.0059H(x; 7, 1)$$
$$+ 0.0023H(x; 8, 1),$$

and

$$F_{W_{K_{6-\max}}}(x) = 0.5322H(x; 1, 1)$$
$$+ 0.2264H(x; 2, 1)$$
$$+ 0.2110H(x; 3, 1)$$
$$+ 0.0004H(x; 5, 1)$$
$$+ 0.0299H(x; 6, 1)$$
$$+ 0.0000H(x; 20, 1).$$

Let $X_{K_{(m+1)-\min(\max)}}$ be a rv with df $F_{X_{K_{(m+1)-\min(\max)}}}(x) = 1 - p + pF_{W_{K_{(m+1)-\min(\max)}}}(x)$ for $x \geq 0$. It follows from Section 4.2 that lower (upper) bounds for the higher-order moments $E[S^j]$ ($j = 4, 5, \ldots$) and the exponential premium principle defined as $\varphi_\eta(S) = (\ln E[e^{\eta S}])/\eta$ ($\eta > 0$) can be found from their counterparts for $X_{K_{(m+1)-\min(\max)}}$. A few numerical values are provided in Tables 5 and 6, respectively.

As expected, the bounds get sharper as the number of moments involved increases.

As for the second approach based on moment bounds with discrete expected stop-loss transforms, Table 7 presents the values of TVaR for $X_{K_{m-down(up)}}$ ($m = 4, 5$) with df $F_{X_{K_{m-down(up)}}}(x) = 1 - p + pF_{W_{K_{m-down(up)}}}(x)$ for $x \geq 0$.

Table 5. Higher-order moments of S, $X_{K_{(m+1)-min}}$ and $X_{K_{(m+1)-max}}$ ($m = 4, 5$)

j	$E[X^j_{K_{5-min}}]$	$E[X^j_{K_{6-min}}]$	$E[S^j]$	$E[X^j_{K_{6-max}}]$	$E[X^j_{K_{5-max}}]$
4	138.7579	138.7579	138.7579	138.7579	138.7579
5	1125.9592	1129.1880	1129.1880	1129.1880	1149.9348
6	10748.5738	10873.8020	10881.2732	10922.7337	11993.6176

Table 6. Exponential premiums of S, $X_{K_{(m+1)-min}}$ and $X_{K_{(m+1)-max}}$ ($m = 4, 5$)

η	$\varphi_\eta(X_{K_{5-min}})$	$\varphi_\eta(X_{K_{6-min}})$	$\varphi_\eta(S)$	$\varphi_\eta(X_{K_{6-max}})$	$\varphi_\eta(X_{K_{5-max}})$
0.2	1.5545	1.5546	1.5546	1.5548	1.5564
0.1	1.3536	1.3536	1.3536	1.3536	1.3536
0.01	1.2137	1.2137	1.2137	1.2137	1.2137

Table 7. Values of TVaR for S, $X_{K_{m-down}}$ and $X_{K_{m-up}}$ ($m = 4, 5$)

	Exact	4 moments		5 moments	
κ	$TVaR_\kappa(S)$	$TVaR_\kappa(X_{K_{4-down}})$	$TVaR_\kappa(X_{K_{4-up}})$	$TVaR_\kappa(X_{K_{5-down}})$	$TVaR_\kappa(X_{K_{5-up}})$
0.9	5.0696	4.9222	5.2062	4.9800	5.1490
0.95	6.2214	5.9708	6.4548	6.0594	6.3642
0.99	8.8460	8.2899	9.3301	8.4655	9.1767
0.995	9.9589	9.2500	10.5629	9.4679	10.3775
0.999	12.5066	11.4122	13.4854	11.7323	13.1382

Table 8. Minimal/Maximal values of VaR within $\mathcal{ME}^{ext}(\mu_m, A_{20}, 1)$ ($m = 4, 5$) vs $VaR_\kappa(S)$

	Exact	$\mathcal{ME}^{ext}(\mu_4, A_{20}, 1)$		$\mathcal{ME}^{ext}(\mu_5, A_{20}, 1)$	
κ	$VaR_\kappa(S)$	$VaR_\kappa(X)_{min}$	$VaR_\kappa(X)_{max}$	$VaR_\kappa(X)_{min}$	$VaR_\kappa(X)_{max}$
0.9	3.3965	3.3896	3.4031	3.3897	3.4001
0.95	4.5704	4.5539	4.6030	4.5584	4.5730
0.99	7.2334	7.2182	7.2690	7.2251	7.2533
0.995	8.3604	8.3143	8.3946	8.3539	8.3925
0.999	10.9388	10.7191	10.9781	10.9226	10.9536

As expected, the inequality (14) is verified. Also, we observe that the interval estimate of $TVaR_\kappa(S)$ shrinks as the number of moments matched increases.

Finally, given that neither method is applicable for the VaR risk measure, we make use of the technique discussed in Remark 10. We find the numerical values provided in Table 8.

For this example, we observe that the exact values of $VaR_\kappa(S)$ are within the minimal and maximal values of their corresponding risk measures among all members of $ME^{ext}(\mu_m, A_{20}, 1)$ for $m = 4, 5$.

Also, the spread between the minimal and maximal values of VaR are reduced when we go from 4 to 5 moments. An identical exercise for the TVaR risk measure resulted in the same conclusions.

References

Abramowitz, M., and I. Stegun, *Handbook of Mathematical Functions,* 9th ed., Washington, DC: US Government Printing Office, 1972.

Altiok, T., "On the Phase-Type Approximations of General Distributions," *IIE Transactions* 17:2, 1985, pp. 110–116.

Bobbio, A., A. Horváth, and M. Telek, "Matching Three Moments with Minimal Acyclic Phase Type Distributions," *Stochastic Models* 21, 2005, pp. 303–326.

Bowers, N.L., H.U. Gerber, G.C. Hickman, D.A. Jones, and C.J. Nesbit, *Actuarial Mathematics,* Schaumburg, Illinois: Society of Actuaries, 1997.

Chaubey, Y.P., J. Garrido, and S. Trudeau, "On the Computation of Aggregate Claims Distributions: Some New Approximations," *Insurance: Mathematics and Economics* 23:3, 1998, pp. 215–230.

Cheung, E.C.K., and J.K. Woo, "On the Discounted Aggregate Claim Costs until Ruin in Dependent Sparre Andersen Risk Processes," *Scandinavian Actuarial Journal* 2016:1, 2016, pp. 63–91.

Comtet, L., *Advanced Combinatorics,* Boston: Dordrecht-Holland, 1974.

Cossette, H., M. Mailhot, and E. Marceau, "T-Var Based Capital Allocation for Multivariate Compound Distributions," *Insurance: Mathematics and Economics* 50:2, 2012, pp. 247–256.

Cossette, H., P. Gaillardetz, and E. Marceau, "Common Mixtures in the Individual Risk Model," *Bulletin de l'Association suisse des actuaires,* 2002, pp. 131–157.

Courtois, C., and M. Denuit, "Moment Bounds on Discrete Expected Stop-Loss Transforms, with Applications," *Methodology and Computing in Applied Probability* 11, 2009, pp. 307–338.

Courtois, C., and M. Denuit, "Bounds on Convex Reliability Functions with Known First Moments," *European Journal of Operational Research* 177, 2007, pp. 365–377.

Courtois, C., M. Denuit, and S. Van Bellegem, "Discrete S-convex Extremal Distributions: Theory and Applications," *Applied Mathematics Letters* 19, 2006, pp. 1367–1377.

Daykin, C.D., T. Pentikäinen, and H. Pesonen, *Practical Risk Theory for Actuaries,* New York: Chapman-Hall, 1994.

Denuit, M., J. Dhaene, M.J. Goovaerts, and R. Kaas, *Actuarial Theory for Dependent Risks—Measures, Orders and Models,* New York: Wiley, 2005.

Denuit, M., and M. Lefèvre, "Some New Classes of Stochastic Order Relations Among Arithmetic Random Variables, with Applications in Actuarial Sciences," *Insurance: Mathematics and Economics* 20, 1997, pp. 197–214.

Denuit, M., C. Lefèvre, and M. Shaked, "The s-convex Orders among Real Random Variables, with Applications," *Mathematical Inequalities and Applications* 1, 1998, pp. 585–613.

Denuit, M., M. Lefèvre, and M. Mesfioui, "On s-convex Stochastic Extrema for Arithmetic Risks," *Insurance: Mathematics and Economics* 25, 1999, pp. 143–155.

Denuit, M., C. Lefèvre, and M. Shaked, "Stochastic Convexity of the Poisson Mixture Model," *Methodology and Computing in Applied Probability* 2:3, 2000, pp. 231–254.

Denuit, M., C. Lefèvre, and S. Utev, "Generalized Stochastic Convexity and Stochastic Orderings of Mixtures," *Probability in the Engineering and Informational Sciences* 13, 1999, pp. 275–291.

De Vylder, F.E., *Advanced Risk Theory.* A Self-Contained Introduction. Editions de l'Université Libre de Bruxelles—Swiss Association of Actuaries, Bruxelles, 1996.

Dufresne, D., "Fitting Combinations of Exponential to Probability Distributions," *Applied Stochastic Models in Business and Industry* 23, 2007, pp. 23–48.

Gerber, H.U., *An Introduction to Mathematical Risk Theory.* S.S. Huebner Foundation. Philadelphia: University of Pennsylvania, 1979.

Hürlimann, W., "Analytical Bounds for Two Value-at-Risk Functionals," *ASTIN Bulletin* 32, 2002, pp. 235–265.

Johnson, M.A., and M.R. Taaffe, "Matching Moments to Phase Distributions: Mixtures of Erlang Distribution of Common Order," *Stochastic Models* 5, 1989, pp. 711–743.

Karlin, S., and W.J. Studden, *Tchebycheff Systems: With Applications in Analysis and Statistics.* New York: Wiley, 1966.

Landriault, D., and G.E. Willmot, "On the Joint Distributions of the Time to Ruin, the Surplus Prior to Ruin and the Deficit at Ruin in the Classical Risk Model," *North American Actuarial Journal* 13:2, 2009, pp. 252–270.

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

Lee, Y.S., and T.K. Lin, "Higher-Order Cornish Fisher Expansion," *Applied Statistics* 41, 1992, pp. 233–240.

Lenart, A., "The Moments of the Gompertz Distribution and Maximum Likelihood Estimation of its Parameters," *Scandinavian Actuarial Journal* 2014, pp. 255–277.

Lindskog, F., and A.J. McNeil, "Common Poisson Shock Models: Applications to Insurance and Credit Risk Modelling," *ASTIN Bulletin* 33:2, 2003, pp. 209–238.

Marceau, E., *Classical Risk Theory and Schmitter's Problems.* PhD Thesis. Université Catholique de Louvain, Louvain-la-Neuve, 1996.

McNeil, A.J., R. Frey, and P. Embrechts, *Quantitative Risk Management.* Princeton University Press, Princeton, 2005.

Melnikov, A., and Y. Romaniuk, "Evaluating the Performance of Gompertz, Makeham and Lee–Carter Mortality Models for Risk Management with Unit-Linked Contracts," *Insurance: Mathematics and Economics* 39, 2006, pp. 310–329.

Milgram, M., "The Generalized Integro-Exponential Function," *Mathematics of Computation* 44:170, 1985, pp. 443–458.

Müller, A., and D. Stoyan, *Comparison Methods for Stochastic Models and Risks,* New York: Wiley, 2002.

Osogami, T., and M. Harchol-Balter, "Closed Form Solutions for Mapping General Distributions to Quasi-Minimal PH Distributions," *Performance Evaluation* 63, 2006, pp. 524–552.

Prékopa, A., "The Discrete Moment Problem and Linear Programming," *Discrete Applied Mathematics* 27, 1990, pp. 235–254.

Ramsay, C., "A Note on the Normal Power Approximation," *ASTIN Bulletin* 21, 1991, pp. 147–150.

Seal, H.L., "Approximation to Risk Theory's F(x,t) by Means of the Gamma Distribution," *ASTIN Bulletin* 9, 1977, pp. 213–218.

Shaked, M., and J. G. Shanthikumar, *Stochastic Orders and their Applications,* 2nd ed., New York: Springer-Verlag, 2007.

Telek, M., and A. Heindl, "Matching Moments for Acyclic Discrete and Continuous Phase-Type Distributions of Second Order," *International Journal of Simulation Systems, Science and Technology* 3:3–4, 2002, pp. 47–57.

Tijms, H. C., *Stochastic Models: An Algorithmic Approach.* Chichester: John Wiley, 1994.

Vanden Bosch, P. M., D. C. Dietz, and E. A. Pohl, "Moment Matching Using a Family of Phase-type Distributions," *Communications in Statistics—Stochastic Models* 16:3–4, 2000, pp. 391–398.

Venter, G., "Transformed Beta and Gamma Distributions and Aggregate Losses" *Proceedings of the Casualty Actuarial Society,* 1983, pp. 156–193.

Whitt, W., "Approximating a Point Process by a Renewal Process, I: Two Basic Methods," *Operation Research* 30:1, 1982, pp. 125–147.

Willmot, G. E., and J. K. Woo, "On the Class of Erlang Mixtures with Risk Theoretic Applications," *North American Actuarial Journal* 11:2, 2007, pp. 99–115.

Willmot, G. E., and X. S. Lin, "Risk Modelling with the Mixed Erlang Distribution," *Applied Stochastic Models in Business and Industry* 27, 2011, pp. 2–16.