CONTAMINATED EXPONENTIAL DISPERSION LOSS MODELS

Zinoviy M. Landsman* and Udi E. Makov†

ABSTRACT

A new family of contaminated exponential dispersion loss models is defined and some of its properties are examined. These models offer a wider family of loss distributions, allowing the modeling of extreme claims. Their usefulness is illustrated with real data.

1. INTRODUCTION

In this paper, we introduce a new statistical model, the *contaminated exponential dispersion loss model* (CEDLM), which is especially designed to handle both ordinary and extremely large losses. The CEDLM is represented as a mixture of two types of loss distributions, where both are assumed to belong to the same *exponential dispersion model* (EDM), having an identical canonical parameter and different *dispersion* parameters. Thus, the CEDLM essentially generalizes the *Tukey model* (Tukey 1960), in which a scale mixture of two normal distributions with identical means and different variances was considered (see also Lindsay 1995).

The introduction of scale mixture of normal distributions left a significant impact on the development of statistical inference, as it allowed the consideration of robustness (see Huber 1981, Chap. 1). It is believed that the CEDLM will offer additional and considerable flexibility in representing fat-tailed[1] distributions, as encountered in the context of insurance and finance.

Tukey referred to the model as a *contaminated model*, in which a "core" distribution was mixed with a "contaminant." It must be emphasized that, while we keep this terminology of "core" and "contaminant" distributions, unlike in the robust approach, we regard claims of both types as relevant for our analysis. While the skewness and kurtosis of a single EDM are determined by the mean, variance function, and dispersion parameter, the CEDLM is flexible enough to possess any degree of skewness and kurtosis. This makes the latter more suitable for fitting models where some of the probability mass is shifted either to the center or to the tails of the distribution.

The aim of this paper is to introduce *contaminated exponential dispersion loss models* and specify their components. This is done in Section 2. In Section 3, a method for estimating the model's parameters is discussed and the properties skewness and kurtosis of the CEDLM are proved. Section 4 compares the CEDLM to the EDM with respect to the limited expected value (LEV). In Section 5, the model is illustrated with real data, and the impact of contamination on LEV, with respect to retention level, is demonstrated.

2. THE CONTAMINATED EXPONENTIAL DISPERSION LOSS MODEL

We define CEDLM as a family of densities

$$\begin{aligned} f_\varepsilon(x|\theta, \lambda, \beta) &= (1-\varepsilon)\exp\{\lambda[\theta x - k(\theta)]\}q_\lambda(x) \\ &\quad + \varepsilon \exp\{\lambda\beta[\theta x - k(\theta)]\}q_{\lambda\beta}(x), \\ &\quad x \in S \subset R, \end{aligned} \tag{1}$$

where $\varepsilon \in (0, 1)$ is a contamination parameter, $\theta \in \Theta$ is a canonical parameter, $\lambda > 0$ is a dispersion parameter, $\beta > 0$ is a dispersion coefficient, and S is a support of the family.

---

* Zinoviy M. Landsman, Ph.D., is a Faculty at Department of Statistics, Senior Researcher at the Actuarial Research Center, University of Haifa, Mount Carmel, Haifa 31905, Israel, e-mail: landsman@stat.haifa.ac.il.

† Udi E. Makov, Ph.D., is Director, Actuarial Programme, Department of Statistics, University of Haifa, Mount Carmel, Haifa 31905, Israel, e-mail: makov@stat.haifa.ac.il.

[1] "Fat-tail" here is synonymous with "thick-tail," which is defined, for example, in Kaas et al. (2001, Sect. 10.3).

As a special case of the CEDLM, we find that, when $\beta = 1$,

$$\begin{aligned} f_\varepsilon(x|\theta, \lambda, 1) &= f(x|\theta, \lambda) \\ &= \exp\{\lambda[\theta x - k(\theta)]\} q_\lambda(x), \end{aligned} \quad (2)$$

which is the EDM. Here, $q_\lambda(x) = dQ_\lambda/dx \geq 0$ is the Radon-Nikodym derivative of the measure Q_λ, which generates EDM, and $k(\theta)$ is the cumulant function, which plays an important role in variance structure of the model. EDM was considered in Tweedie (1984), Nelder and Wedderburn (1972), Nelder and Verrall (1997), and Jorgensen (1986, 1987, 1992, and 1997). Current interest in the EDM is due to Jorgensen, who outlines the EDM as one of the main classes of dispersion models, which includes most standard distribution families.

In Landsman and Makov (1998; 1999a,b; and 2000), the insurance (credibility) aspects of the EDMs were discussed, and the intrinsic relationship between linear Bayes estimators and stochastic approximation was established.

The EDM has certain analogies with location and scale models, where location is expressed by the population mean

$$\mathrm{E}X = k'(\theta) = \mu. \quad (3)$$

The role of the scale parameter is played by the so-called *dispersion parameter* $\sigma^2 = 1/\lambda$. It follows from Equation (2) that the population variance is given by

$$\mathrm{Var}X = k''(\theta)/\lambda = V(\mu)/\lambda, \quad (4)$$

where $V(\mu)$ is called the *variance function.*

The variance function plays an important role in the classification of the EDM. Morris (1982 and 1983) studied exponential families with quadratic variance functions. Cubic variance functions were discussed by Mora (1986) and Letac and Mora (1990). Power variance functions outline Tweedie's exponential dispersion class (Tweedie 1984, Bar-Lev and Enis 1986, and Bar-Lev et al. 1992). In Table 1, we give some examples of the EDM.

At the root of the CEDLM is the assumption that $[(1 - \varepsilon)100\%]$ of the claims or losses follow the core loss distribution $f(x|\theta, \lambda)$. A percentage of the losses ($\varepsilon 100\%$) does not follow the core distribution, constituting very large losses unaccounted for by this distribution, mainly due to the fact that its tail is not sufficiently thick.

The distribution of these losses is given by $f(x|\theta, \lambda\beta)$. The resulting four-parameter $(\theta, \lambda, \beta, \varepsilon)$ loss distribution is very rich and can allow the modeling of fat-tailed loss data, as will be shown in Section 3. Both densities in the CEDLM have a common canonical parameter θ and cumulant function $k(\theta)$ and different dispersion parameters; therefore, from Equations (3) and (4), it follows that they have a common expectation and different variances. The variance ratio of the contaminant to the core density is $1/\beta$.

Note that the CEDLM is a mixture, with respect to the variance parameter, of two exponential dispersion distributions, and, therefore, is essentially different from the traditional mixtures suggested for the exponential family, where the mixing, in a Bayesian fashion, is done with

Table 1

Examples of the EDM

Distributions	Variance Function $V(\mu)$	Class	$k(\theta)$
Normal	1	Morris, Tweedie	$\frac{1}{2}\theta^2$
Poisson	μ	Morris, Tweedie	$\exp(\theta)$
Compound Poisson	$\mu^p,\ 1 < p < 2$	Tweedie	$\frac{1}{2-p}\left(\frac{\theta}{1-p}\right)^p,\ \theta < 0$
Bernoulli	$\mu(1 - \mu)$	Morris	$\ln(1 + \exp(\theta))$
Negative Binomial	$\mu(1 + \mu)$	Morris	$-\ln(1 - \exp(\theta))$
Gamma	μ^2	Morris, Tweedie	$-\ln(-\theta),\ \theta < 0$
Hyperbolic sectant	$1 + \mu^2$	Morris	$-\ln(\cos(\theta)),\ \lvert\theta\rvert < \frac{\pi}{2}$
Inverse Gauss	μ^3	Tweedie, Letac	$-\sqrt{-2\theta},\ \theta \leq 0$
Hyperbolic	$(1 + \mu^2)^{3/2}$	Letac	$-\sqrt{1 - \theta^2},\ \lvert\theta\rvert \leq 1$

respect to the canonical parameter (see, e.g., Lindsay 1995).

The mixture model proposed here has a universal population mean for both components. In certain actuarial contexts, such a situation arises naturally, that is, two subpopulations characterized by the same mean and different degree of variability.[2]

3. Some Properties of the CEDLM

In this section, we investigate the proposed CEDLM. In particular, it is shown that, by varying the parameter β, the model's skewness and kurtosis can be fine-tuned and increased (or decreased) as required by the loss data. The structure of Equation (1) implies that moments of the loss X are a mixture of moments of the EDM. This simplifies the evaluation of moments of the CEDLM. It is straightforward to verify that the second and third moments around the mean of Equation (1) are, respectively,

$$M_2(X) = V_{\mu,\lambda,\varepsilon,\beta}(X) = \frac{1}{\lambda} V(\mu)\left(1 - \varepsilon + \frac{\varepsilon}{\beta}\right)$$

$$\begin{aligned} M_3(X) &= \mathrm{E}_{\mu,\lambda,\varepsilon,\beta}(X - \mu)^3 \\ &= \frac{1}{\lambda^2} V(\mu)V'(\mu)\left(1 - \varepsilon + \frac{\varepsilon}{\beta^2}\right), \end{aligned} \tag{5}$$

where prime indicates a first derivative.

A crucial parameter in modeling losses is the degree of skewness, which indicates the ability of the model to account for large claims or losses. We now establish the skewness of the CEDLM and the impact of the variance ratio, β, on it.

Theorem 1

For any degree of contamination $\varepsilon \in (0, 1)$, the skewness of the core EDM can be increased as much as desired.

Proof

Recalling that the degree of skewness, γ_1, is defined as $M_3(X)/[M_2(X)]^{3/2}$, we obtain from Equation (5)

$$\gamma_1(\lambda, \varepsilon, \beta) = \gamma_1(\lambda) \frac{\left(1 - \varepsilon + \dfrac{\varepsilon}{\beta^2}\right)}{\left(1 - \varepsilon + \dfrac{\varepsilon}{\beta}\right)^{3/2}}, \tag{6}$$

where $\gamma_1(\lambda, \varepsilon, \beta)$ is the skewness of the CEDLM and

$$\gamma_1(\lambda) = \frac{1}{\sqrt{\lambda}} \frac{V'(\mu)}{\sqrt{V(\mu)}}$$

is the skewness of the "core" EDM with dispersion parameter λ. As the derivative of the function

$$\varphi(\beta) = \frac{\left(1 - \varepsilon + \dfrac{\varepsilon}{\beta^2}\right)}{\left(1 - \varepsilon + \dfrac{\varepsilon}{\beta}\right)^{3/2}} \tag{7}$$

is equal to

$$\varphi'(\beta) = \frac{3\varepsilon}{2\beta^4} \frac{(1 - \varepsilon)\beta^2 - 4/3(1 - \varepsilon)\beta - \varepsilon/3}{(\varepsilon/\beta + (1 - \varepsilon))^{5/2}},$$

$$\beta \neq 0$$

and reduces to 0 only at points

$$(2/3)(1 - \sqrt{1 + (3/4)\varepsilon/(1 - \varepsilon)}),$$

and

$$(2/3)(1 + \sqrt{1 + (3/4)\varepsilon/(1 - \varepsilon)}),$$

one can see that, for any ε, the interval

$$(0, 1] \subset \left[\frac{2}{3}\left(1 - \sqrt{1 + \frac{3}{4}\frac{\varepsilon}{1 - \varepsilon}}\right), \frac{2}{3}\left(1 + \sqrt{1 + \frac{3}{4}\frac{\varepsilon}{1 - \varepsilon}}\right)\right],$$

and so $\varphi'(\beta) < 0$ if $\beta \in (0, 1]$. Hence, $\varphi(\beta)$ decreases monotonically in (0, 1]. As to the end points, they are $\varphi(0) = \infty$, $\varphi(1) = 1$. ☐

Remark 1

This result is quite intuitive. As β decreases, the variance of the contaminated loss distribution increases relative to the variance of the core loss distribution. This allows for losses bigger than

---

[2] Our study is confined to mixing EDMs having a common mean. When the constraint on a common mean is removed, which is a possibility, the number of parameters in the model increases. In general, one may consider a mixture of distributions not belonging to the EDM, like the lognormal. Such generalizations, which essentially complicate the model, will not be discussed here.

those accounted for by the core distribution and, hence, a more skewed model. Note that, from Equations (6) and (7), it follows that the function-multiplier $\varphi(\beta)$, which increases the skewness of the core distribution, depends only on ε and β and does not depend on the characteristics of the EDM; that is, $\varphi(\beta)$ is model-free and is the same for all members of the EDM. For example, let $\varepsilon = 0.2$ and suppose that one wishes to triple the skewness of the core. One should simply solve the equation $\varphi(\beta) = 3$ and obtain (numerically) that $\beta = 0.156$. In Figure 1, we show the graph of the core gamma distribution and a contaminated gamma whose skewness is tripled.

The following theorem establishes the impact of the variance ratio β on the kurtosis of the CEDLM, which can have bearing on the concentration of probability mass in the tails.

Theorem 2

For any degree of contamination $\varepsilon \in (0, 1)$, the kurtosis of the core EDM,

$$\gamma_2(\lambda) = \frac{1}{\lambda}\left(V''(\mu) + \frac{V'(\mu)}{V(\mu)}\right) + 3,$$

can be increased as much as desired, with β decreased from 1 to 0.

PROOF

Straightforward calculation provides the fourth central moment:

$$M_4(X) = \mathrm{E}_{\mu,\lambda,\varepsilon,\beta}(X - \mu)^4$$

$$= \left(\frac{1}{\lambda^3} V(\mu)(V''(\mu)V(\mu) + V'(\mu)^2) + \frac{3}{\lambda^2} V^2(\mu)\right) \times \left(1 - \varepsilon + \frac{\varepsilon}{\beta^3}\right), \qquad (8)$$

which, together with the formula for the second central moment $M_2(X)$ (see Equation 5), results in an expression for the kurtosis

$$\gamma_2(\lambda, \varepsilon, \beta) = \frac{M_4(X)}{(M_2(X))^2} = \gamma_2(\lambda)\psi(\beta), \qquad (9)$$

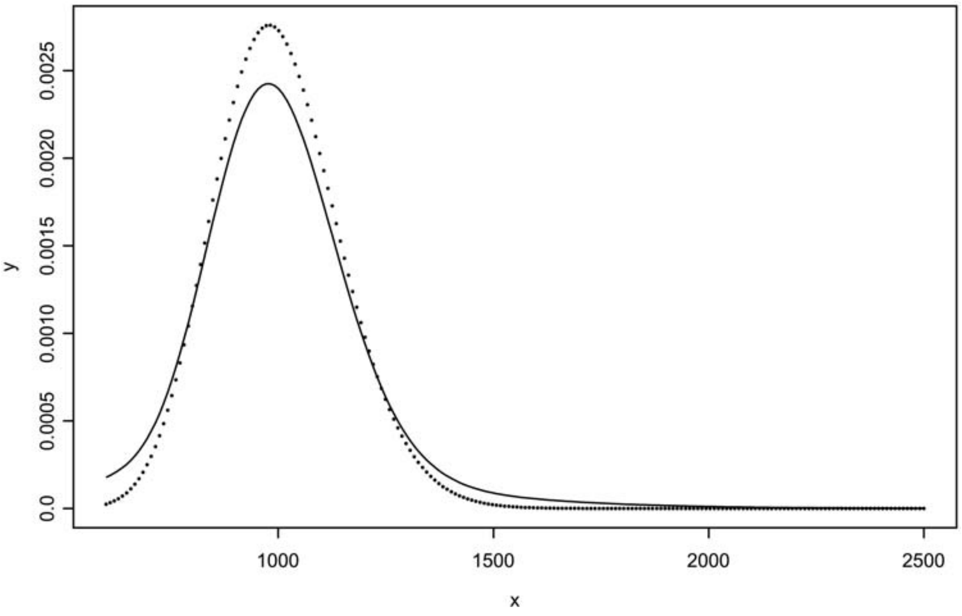
where

$$\gamma_2(\lambda) = \frac{1}{\lambda}\left(V''(\mu) + \frac{V'(\mu)}{V(\mu)}\right) + 3$$

is the kurtosis of the "core" EDM with a dispersion parameter λ and

$$\psi(\beta) = \frac{\left(1 - \varepsilon + \dfrac{\varepsilon}{\beta^3}\right)}{\left(1 - \varepsilon + \dfrac{\varepsilon}{\beta}\right)^2}. \qquad (10)$$

Figure 1
Gamma (Dotted Line) and Contaminated Gamma Distribution



The derivative of $\psi(\beta)$ is equal to

$$\psi'(\beta) = \varepsilon(2(1-\varepsilon)\beta^5(1-\varepsilon+\varepsilon/\beta)^3)^{-1}\psi_1(\beta),$$

where

$$\psi_1(\beta) = \beta^3 - \frac{3}{2}\beta - \frac{\varepsilon}{2(1-\varepsilon)}.$$

It follows, by simple analysis, that the function $\psi_1(\beta)$ has only one extremal point on the line $[0, \infty)$, which is the minimum point

$$\beta_{\min} = \frac{\sqrt{3}}{2}, \quad \varphi_1(\beta_{\min}) = -3\frac{\sqrt{3}}{8} - \frac{\varepsilon}{2(1-\varepsilon)} < 0, \qquad (11)$$

and, further, $\psi_1(0) = -(\varepsilon/2(1-\varepsilon)) < 0$ and $\psi_1(1) = -(1/2) - (\varepsilon/2(1-\varepsilon)) < 0$. We, therefore, conclude that the function $\psi_1(\beta)$ and the function $\psi'(\beta)$ are both negative on the whole interval $[0, 1]$. This implies that $\psi(\beta)$ decreases monotonically in $(0, 1]$. As to the end points, they are $\psi(0) = \infty$ and $\psi(1) = 1$. □

Remark 2

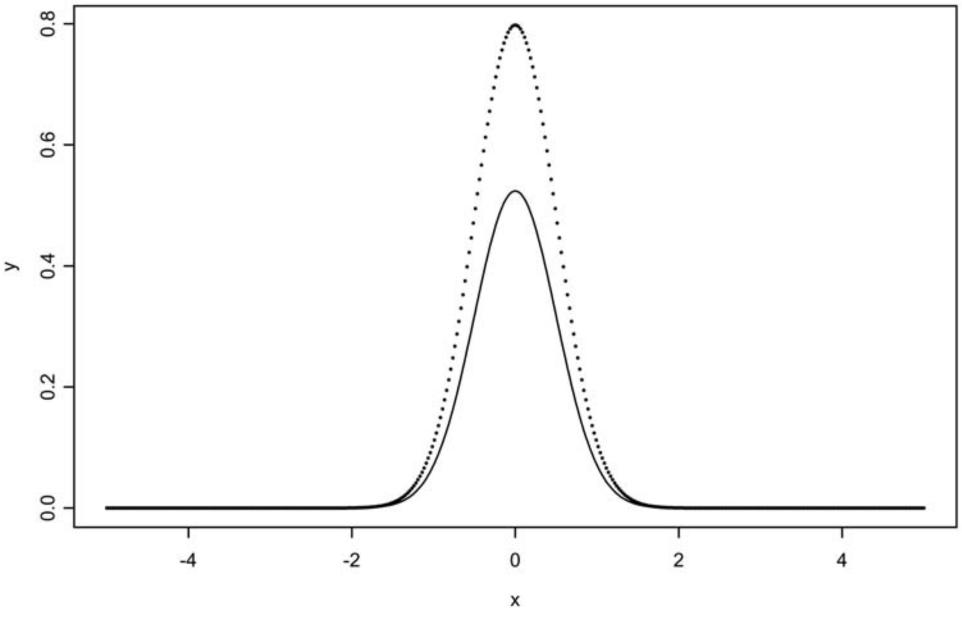
The function-multiplier $\psi(\beta)$, which increases the kurtosis of the core distribution (see Equation 10), depends only on ε and β and does not depend on the characteristics of the EDM; that is, $\psi(\beta)$, like in the skewness case, is model-free. For example, let $\varepsilon = 0.2$ and suppose that one wishes to triple the kurtosis of the core. One should simply solve the equation $\psi(\beta) = 3$ and obtain (numerically) that $\beta = 0.3436$.

Remark 3

The interpretation of kurtosis has been discussed by several authors. Finecan (1964) and Moors (1986) showed that kurtosis may be interpreted as measuring the shift of probability mass away from $\mu \pm \sigma$ toward either the middle of the distribution or the tails, the latter case giving rise to fat-tailed distributions (see also Balanda and MacGillivray 1988 for a useful discussion). Moreover, the kurtosis $\gamma_2(F)$ maintains a stochastic ordering of distribution in the sense that $F < G$ if $\gamma_2(F) \leq \gamma_2(G)$. From Theorem 2 it follows that CEDLM with fixed ε and $\beta \downarrow 0$ can be considered an ordered set of densities directed to probability mass shifting. Decreasing β to 0 could result in a more fat-tailed distribution, regardless of whether the distribution is symmetrical or not.

In Figure 2, we give the graph of the core normal distribution (dotted line) and contaminated

Figure 2
Normal (Dotted Line) and Contaminated Normal Densities



normal (solid line) with $\varepsilon = 0.2$, having zero skewness and tripled kurtosis; that is, $\beta = 0.3436$ (see Remark 2).

Figure 3 shows the graph of the core inverse Gaussian model and contaminated inverse Gaussian model (CIGM), whose density function is

$$\begin{aligned} f_\varepsilon(x|\theta, \lambda, \beta) = (1-\varepsilon) \\ \times \exp(\lambda(x\theta + \sqrt{(-2\theta)})q_\lambda(x) \\ + \varepsilon \exp(\lambda\beta(x\theta + \sqrt{(-2\theta)})q_{\lambda\beta}(x), \quad (12) \end{aligned}$$

where

$$\theta \in (-\infty, 0], q_\lambda(x) = \sqrt{\lambda/(2\pi x^3)} \exp(-\lambda/(2x)),$$

and $x \in (0, \infty)$, with the same parameters of contamination ($\varepsilon = 0.2$, $\beta = 0.3436$) as in the normal case.

Remark 4

Values of β, $\beta \geq \beta_0 > 1$, can also increase the kurtosis of the "core" $\gamma_2(\lambda)$ up to $1/(1-\varepsilon)$ times $\gamma_2(\lambda)$. As shown below, if, without loss of generality, $\varepsilon \in (0, 0.5]$, then β_0 can be chosen to be 2.5469. In fact, Equation (11) indicates that $\psi_1(\beta)$ has only one positive root, and, if $0 < \varepsilon \leq 0.5$; that is, $0 < \varepsilon/(1-\varepsilon) \leq 1$, then

$$\psi_1(\beta) \geq \beta^3 - \tfrac{3}{2}\beta - \tfrac{1}{2} \geq 0,$$

if $\beta \geq \beta_1$, where β_1 is a positive solution of the cubic equation $\beta^3 - (3/2)\beta - 1/2 = 0$. This implies that $\psi(\beta)$ increases for $\beta \geq \beta_1$ and, moreover, $\psi(\beta) \uparrow 1/(1-\varepsilon)$ if $\beta \uparrow \infty$. One can show straightforwardly that condition $\psi(\beta) \geq 1$ is equivalent to $\beta^3 - 2\beta^2 - (\varepsilon/(1-\varepsilon))\beta + 1 \geq 0$. If $\varepsilon/(1-\varepsilon) \leq 1$. Therefore, one can have

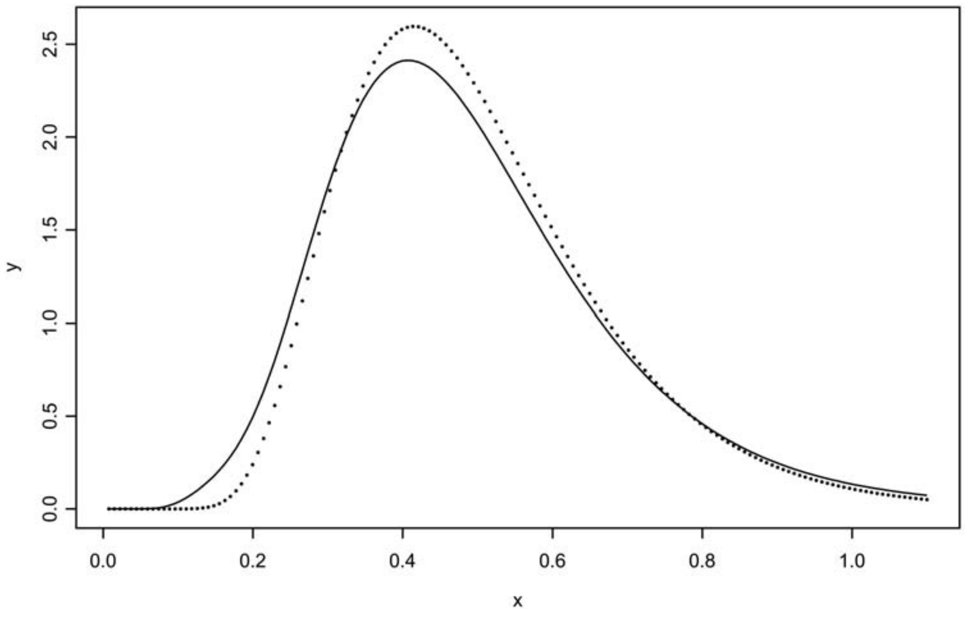
$$\begin{aligned} \beta^3 - 2\beta^2 \\ - \frac{\varepsilon}{(1-\varepsilon)}\beta + 1 \geq \beta^3 - 2\beta^2 - \beta + 1 \geq 0, \end{aligned}$$

which holds if $\beta \geq \beta_2$, where β_2 is the positive root of equation $\beta^3 - 2\beta^2 - \beta + 1 = 0$. Then one can take $\beta_0 = \max(\beta_1, \beta_2)$. Numerical study shows that $\beta_1 \leq 1.36603$ and $\beta_2 \leq 2.5469$, so one can chose $\beta_0 = 2.5469$.

Fitting the CEDLM requires estimating the model's parameters. We now discuss several moment estimators. As the "core" and "contaminant" have the same expectation, μ, the sample mean, $\bar{x} = 1/n \sum_{i=1}^{n} x_i$, is a consistent estimator of μ. When all three parameters ε, λ and β are unknown, estimation of the mixture parameters often becomes extremely tedious. The formula for the fourth central moment (8) together with Equation (5) allows the following system of equations:

Figure 3

Inverse Gauss (Dotted Line) and Contaminated Inverse Gauss Densities



$$\begin{cases} \dfrac{1}{\lambda} V(\bar{x})\left(1-\varepsilon+\dfrac{\varepsilon}{\beta}\right)=s^2 \\ \dfrac{1}{\lambda^2} V(\bar{x})V'(\bar{x})\left(1-\varepsilon+\dfrac{\varepsilon}{\beta^2}\right)=m_3 \\ \left(\dfrac{1}{\lambda^3} V(\bar{x})(V''(\bar{x})V(\bar{x})+V'(\bar{x})^2)+\dfrac{3}{\lambda^2} V^2(\bar{x})\right) \\ \qquad \times\left(1-\varepsilon+\dfrac{\varepsilon}{\beta^3}\right)=m_4 \end{cases} \tag{13}$$

Here, $s^2 = (1/n)\sum_{i=1}^{n}(x_i-\bar{x})^2$ and $m_j = (1/n)\sum_{i=1}^{n}(x_i-\bar{x})^j, j=3,4$ are third and fourth central sample moments, respectively. When λ is assumed known, the first two equations in Equation (13) give explicit simple moment estimators for the CEDLM with non-Gaussian "core" distribution. For the case where λ is unknown, the third equation in Equation (13) is required and a solution to the system of equations can be obtained by iteration procedures. Such an extension, as well as the application of iterative maximum likelihood techniques, will not be discussed.

Theorem 3

Suppose $V'(\mu) \neq 0$, and λ is known, then the moment estimators of parameters μ, β, and ε are

$$\hat{\mu} = \bar{x}$$

$$\hat{\beta} = \frac{\dfrac{1}{\lambda} V'(\bar{x})\left(s^2 - \dfrac{1}{\lambda} V(\bar{x})\right)}{m_3 - \dfrac{1}{\lambda} V'(\bar{x})s^2}$$

$$\hat{\varepsilon} = \frac{\left(s^2 - \dfrac{1}{\lambda} V(\bar{x})\right)^2}{m_3 - 2\dfrac{1}{\lambda} V'(\bar{x})s^2 + \dfrac{1}{\lambda^2} V'(\bar{x})V(\bar{x})} \frac{V'(\bar{x})}{V(\bar{x})},$$

where s^2 and m_3 are, respectively, the sample variance and sample third central moment. In addition, $\hat{\varepsilon} \in [0, 1]$ if

$$\frac{m_3}{s^4} \geq \frac{V'(\bar{x})}{V(\bar{x})} \tag{14}$$

PROOF

From Equation (5), after replacing μ, $M_2(X)$ and $M_3(X)$ with $\bar{x}$, s^2, and m_3, respectively, and solving this system with respect to β and ε, one obtains that

$$\hat{\beta} = \frac{1-\lambda D_1}{\lambda(D_1 - \lambda D_2)},$$

$$\hat{\varepsilon} = \frac{(1-\lambda D_1)^2}{(1-\lambda D_1)^2 + \lambda^2(D_2 - D_1^2)}, \tag{15}$$

where

$$D_1 = \frac{s^2}{V(\bar{x})},$$

$$D_2 = \frac{m_3}{V(\bar{x})V'(\bar{x})}.$$

Condition (14) follows immediately from Equation (15). □

4. COMPARISON OF THE CEDLM AND THE EDM WITH RESPECT TO THE LIMITED EXPECTED VALUE

We compare the contaminated distribution with its "core" with respect to the limited expected value (LEV), which is essential when considering reinsurance and deductible discounts.

LEV at level $R > 0$ of loss value X having distribution function $F(x)$ is defined as

$$LEV_F(R) = \mathrm{E}(\min(X, R)) = \int_0^R (1 - F(x))\, dx.$$

In excess of loss reinsurance with retention level M, $LEV_F(M)$ represents the expected amount of claims paid by the insurance company if all claims above M are dealt with by the reinsurance company. In deductible discounts, $LEV_F(\infty) - LEV_F(D)$ represents the expected amount of payment made by an insurance company if the insured bears a deductible cost D. In this context, the deductible discount coefficient (DDC) is defined as

$$DDC_F(D) = 1 - \frac{LEV_F(D)}{\mathrm{E}(X)}. \tag{16}$$

(For further details, see Klugman et al. 1998 and Daykin et al. 1994).

Note that the distribution function of the contaminated observation follows the same mixture structure of the CEDLM; that is,

$$F_{\varepsilon,\lambda,\beta}(x) = (1-\varepsilon)F_\lambda(x) + \varepsilon F_{\lambda\beta}(x),\ x \in R, \tag{17}$$

where $F_{\lambda}(x)$ is an EDM with dispersion parameter λ. Consequently, the LEV for the CEDLM takes the form

$$LEV_{F_{\varepsilon,\lambda,\beta}}(M) = (1 - \varepsilon)LEV_{F_{\lambda}}(M) + \varepsilon LEV_{F_{\lambda\beta}}(M). \quad (18)$$

To compare the contaminated model with its "core," we suggest calculating the ratio of their LEVs and DDCs; that is,

$$R_1(M) = \frac{LEV_{F_{\varepsilon,\lambda,\beta}}(M)}{LEV_{F_{\lambda}}(M)} = (1 - \varepsilon) + \varepsilon \frac{LEV_{F_{\lambda\beta}}(M)}{LEV_{F_{\lambda}}(M)} \quad (19)$$

and

$$\begin{aligned} R_2(D) &= \frac{DDC_{F_{\varepsilon,\lambda,\beta}}(D)}{DDC_{F_{\lambda}}(D)} \\ &= (1 - \varepsilon) + \varepsilon \frac{\mu - LEV_{F_{\lambda\beta}}(D)}{\mu - LEV_{F_{\lambda}}(D)}. \end{aligned} \quad (20)$$

In the next section, we fit our model to real data and show that the coefficients $R_1(M)$ and $R_2(D)$ can differ from 1 considerably. The implications of this also are discussed.

5. ILLUSTRATION USING REAL DATA

5.1 Data Modeling with CEDLM

We illustrate the model by using the data in Table 2 of U.K. fire losses, provided by Klugman et al. (1998). We fit a CEDLM to the data and compare the fit to that of other common models. This section aims to provide an illustration only and is not intended as a comparative numerical study of various models. The particular data we analyzed exhibit two characteristics often found in loss data, namely, a large peak at the low end of the distribution and an extended tail due to extremely large claims.

Table 3

Minimal Value of χ^2

Model	χ^2
Gamma	12,645
Inverse Gaussian	908
Lognormal	95
Transformed beta	41.57

As the losses are given in grouped form, we choose to minimize the χ^2 distance for fitting an appropriate model. Table 3 compares the goodness of fit of the gamma, inverse Gaussian, lognormal, and transformed beta models, as obtained by minimizing the χ^2 distance.

One can see that none of the models considered can adequately fit this data. We propose, instead, to consider a CEDLM and, in particular, the CIGM, whose density function was given in Section 3 (see Equation 12). It transpires that this model gives a dramatically better result. The best fit of the CIGM with $\mu = 7008$, $\lambda = 46.95$, $\varepsilon = 0.487$, and $\beta = 13.236$ resulted in $\chi^2 = 21.27$, which testifies to the quality of the fit. (This is supported by $P_{value} = 0.4353$). The advantage of the CIGM is indicated not only by the reduced value of χ^2, but also through the use of Q-Q plots (see Figures 4–6). These plots illustrate the superior fit of the CIGM to the U.K. fire loss data as compared to that of the transformed beta or lognormal models.

The graph of the histogram of the data and the fitted density of the CIGM are given in Figure 7. Here, the dotted line corresponds to the "core"

Table 2

U.K. Fire Losses

Upper Limit	Number	Upper Limit	Number	Upper Limit	Number
100	4,319	2,260	805	51,200	136
140	795	3,200	694	72,410	108
200	910	4,530	602	102,400	88
280	962	6,400	480	250,000	117
400	1,097	9,050	382	500,000	47
570	1,121	12,800	329	750,000	12
800	1,046	18,100	273	1,000,000	4
1,130	969	25,600	214	2,000,000	8
1,600	843	36,200	172	3,000,000	3

Source: Klugman et al. (1998).

Figure 4

Lognormal Q-Q Plot for U.K. Fire Losses

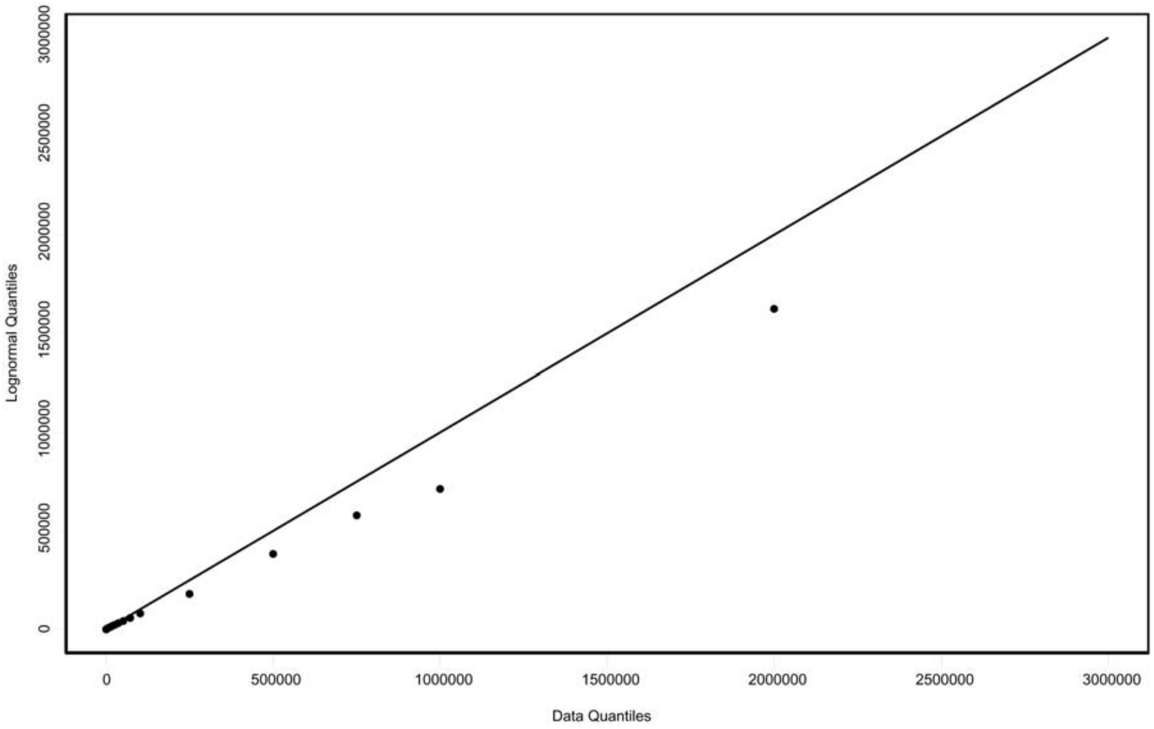


Figure 5

Transformed Beta Q-Q Plot for U.K. Fire Losses

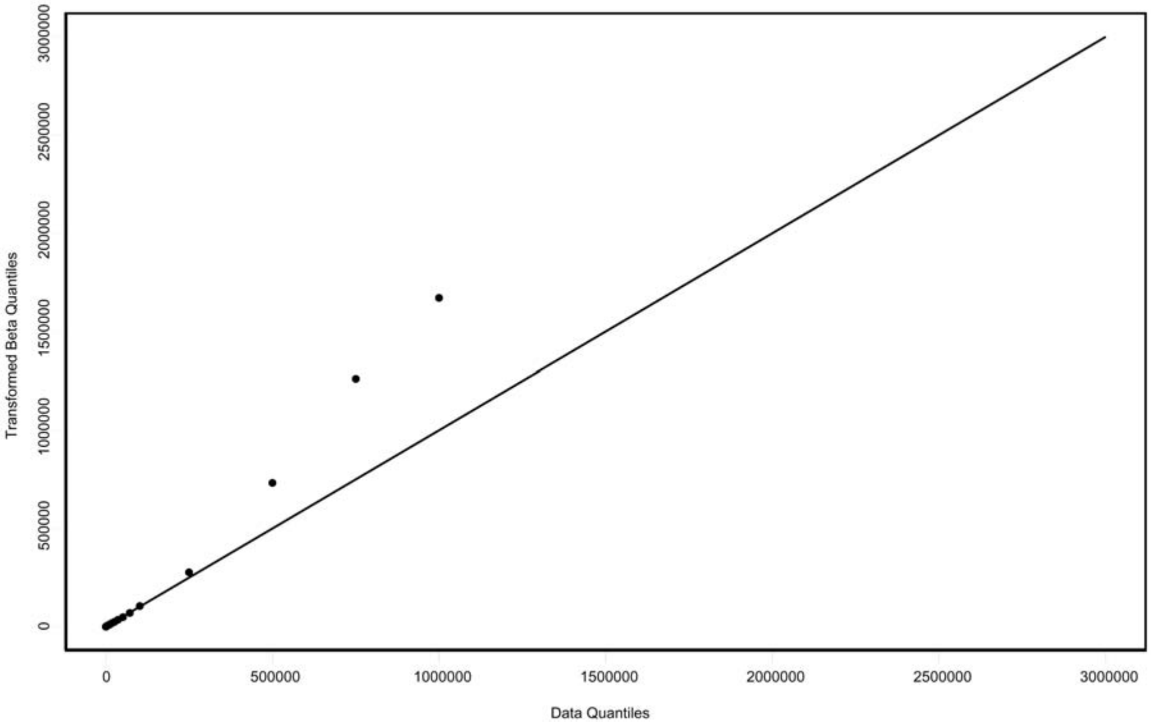


Figure 6

Contaminated Inverse Gauss Q-Q Plot for U.K. Fire Losses

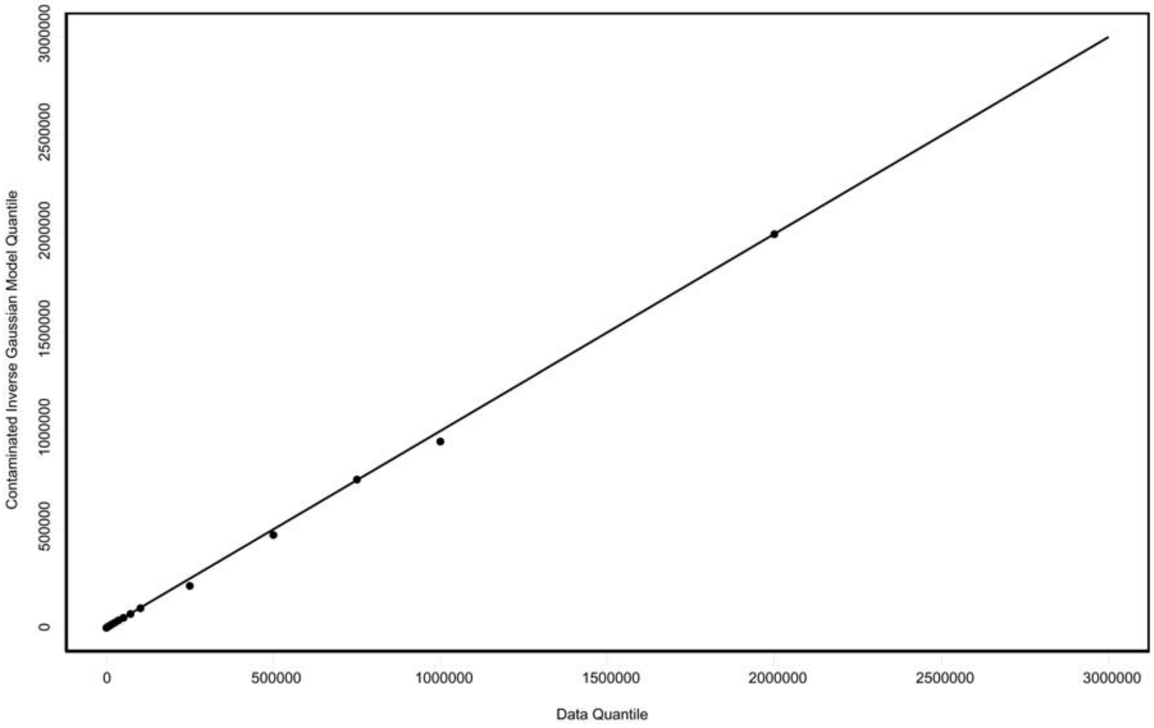
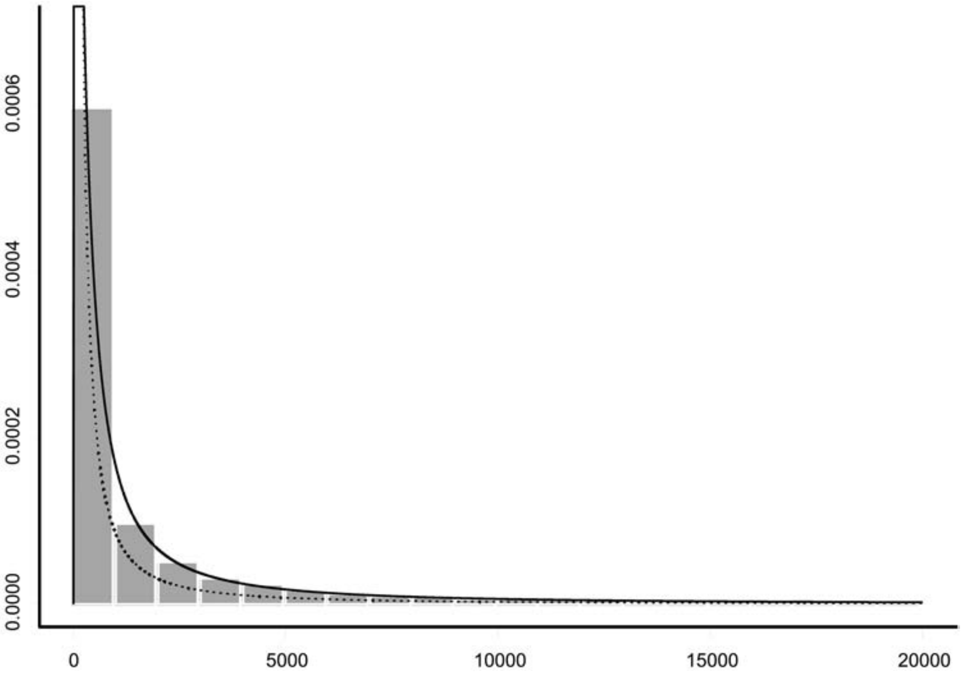


Figure 7

Histogram of U.K. Fire Losses. Best Fitted CIGM and Its Core (Dotted Line)



inverse Gaussian distribution (namely, $\varepsilon = 0$ or $\beta = 1$), and the solid line corresponds to the CIGM.

5.2 Impact of Contamination on LEV and DDC

In Figures 3 and 7, we can see that the graphs of the contaminated and "core" inverse Gaussian distributions are not dramatically different. However, their LEV for retention levels common in reinsurance, and their DDC for deductible values acceptable in insurance are significantly different. To show this, let us calculate the ratios R_1 and R_2 considered in the previous section. From Equation (19) and the expression for LEV for inverse Gaussian distribution given in Klugman et al. (1998, A.3.2), we write

$$R_1(M) = (1 - \varepsilon)$$

$$+ \varepsilon \frac{1 - (1 - \mu/M)\Phi((M/\mu - 1)(\lambda/M)^{1/2}\beta^{1/2}) -}{1 - (1 - \mu/M)\Phi((M/\mu - 1)(\lambda/M)^{1/2}) -}$$

$$\frac{- (1 + \mu/M) \exp(2/M\beta)\Phi(-(M/\mu + 1)(\lambda/M)^{1/2}\beta^{1/2})}{- (1 + \mu/M) \exp(2\lambda/M)\Phi(-(M/\mu + 1)(\lambda/M)^{1/2})}$$

$R_2(D)$ can be calculated from Equation (20) and the previous formula.

In Figures 8 and 9, we show the graphs of $R_1(M)$, where M increases from \$100,000 to \$1,000,000, and of $R_2(D)$, where D increases up to 7,000. In both cases, we use the estimated values of the parameters of the CIGM. Both of these figures clearly demonstrate the benefits of using the CEDLM when it fits the data well.

Figure 8

Changing $R_1(M)$ with Reinsurence Level M

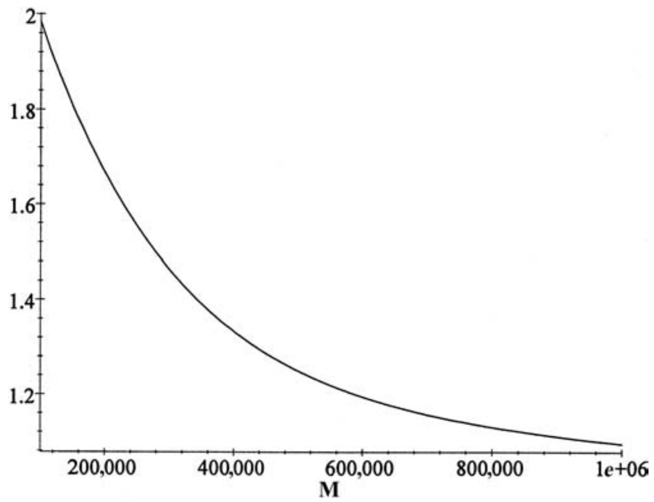
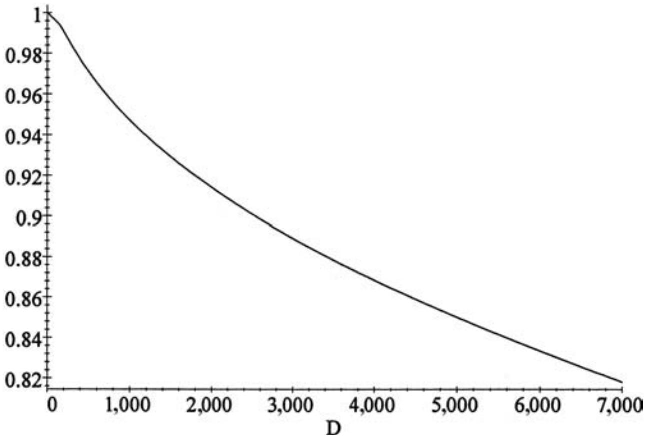


Figure 9

Changing $R_2(D)$ with Deductable Level D



For retention level M = \$400,000, R_1 is approximately 1.33. This implies that the use of an EDM would underestimate the correct expected losses per claimant by about 25%. This is very damaging indeed for an insurance company.

As for deductible discounts, suppose D = \$1,000. R_2 is approximately equal to 0.947 and, hence, the expected losses of the insurance company are actually smaller than those calculated using the EDM. This reduction can justify a reduction in premiums and, hence, an increase in competitiveness.

6. SUMMARY

Whatever the underlying model, large claims are likely to disrupt the modeling process. This is also true for the EDM which offers a rich family of distributions. By modeling a contaminated version of this family, as proposed here, very large observations can be accommodated in the model and, hence, reliable statistical inference is also possible for fat-tailed losses. The benefits from using the CEDLM are obvious. A better fit to the data allows the insurance company to evaluate its premiums more accurately. This clearly has an impact on its cash flow and competitiveness.

ACKNOWLEDGMENT

We are grateful to the Committee on Knowledge Extension Research for its financial support.

REFERENCES

BALANDA, K. P., AND H. L. MACGILLIVRAY. 1988. "Kurtosis: A Critical Review," *American Statistician* 42: 111–19.

BAR-LEV, S. K., AND P. ENIS. 1986. "Reproducibility and Natural Exponential Families with Power Variance Functions," *The Annals of Statistics* 14: 1507–22.

BAR-LEV, S. K., D. BSHOUTY, AND P. ENIS. 1992. "On Polynomial Variance Functions," *Probability Theory and Related Fields* 94: 69–82.

DAYKIN, C. D., T. PENTIKAINEN, AND M. PESONEN. 1994. *Practical Risk Theory for Actuaries.* London: Chapman and Hall.

FINECAN, H. M. 1964. "A Note on Kurtosis," *Journal of the Royal Statistical Society* Ser. B 26: 111–12.

HUBER, P. 1981. *Robust Statistics.* New York: Wiley.

JORGENSEN, B. 1986. "Some Properties of Exponential Dispersion Models," *Scandinavian Journal of Statistics* 13: 187–98.

———. 1987. "Exponential Dispersion Models (with Discussion)." *Journal of the Royal Statistical Society* Ser. B 49: 127–62.

———. 1992. "Exponential Dispersion Models and Extensions: A Review," *International Statistical Review* 60: 5–20.

———. 1997. *The Theory of Dispersion Models.* London: Chapman and Hall.

KAAS, R., M. GOOVAERTS, J. DHAENE, AND M. DENUIT. 2001. *Modern Actuarial Risk Theory.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

KLUGMAN, S. A., H. H. PANJER, AND G. E. WILLMOT. 1998. *Loss Models from Data to Decisions.* New York: Wiley.

LANDSMAN, Z., AND U. MAKOV. 1998. "Exponential Dispersion Models and Credibility," *Scandinavian Actuarial Journal* (1): 89–96.

———. 1999a. "Credibility Evaluation for the Exponential Dispersion Family," *Insurance: Mathematics & Economics* 24: 33–9.

———. 1999b. "On Stochastic Approximation and Credibility," *Scandinavian Actuarial Journal* (1): 15–31.

———. 2000. "On Credibility Evaluation and the Tail Area of the Exponential Dispersion Family," *Insurance: Mathematics & Economics* 27: 277–83.

LETAC, G., AND M. MORA. 1990. "Natural Real Exponential Families with Cubic Variance Functions," *The Annals of Statistics* 18: 1–37.

LINDSAY, B. G. 1995. *Mixture Models: Theory, Geometry and Applications. NSF-CMBS Regional Conference Series in Probability and Statistics,* Vol. 5. Institute of Mathematical Statistics, Hayward, Calif.

MOORS, J. J. A. 1986. "The Meaning of Kurtosis: Darlington Reexamined," *American Statistician* 40: 283–84.

MORA, M. 1986. "Classification des Fonctions Variance Cubiques des Families Exponentielles sur R," Computes Rendusde l'Academie des Sciences: Series 1: Mathematique 302: 587–90.

MORRIS, C. N. 1982. "Natural Exponential Families with Quadratic Variance Functions," *The Annals of Statistics* 10: 65–80.

———. 1983. "Natural Exponential Families with Quadratic Variance Functions: Statistical Theory," *The Annals of Statistics* 11: 515–29.

NELDER, J. A., AND R. W. M. WEDDERBURN. 1972. "Generalized Linear Models," *Journal of the Royal Statistical Society* Ser. A 135: 370–384.

NELDER, J. A., AND R. J. VERRALL. 1997. "Credibility Theory and Generalized Linear Models," *ASTIN Bulletin* 27(3): 71–82.

TWEEDIE, M. C. K. 1984. "An Index Which Distinguishes between Some Important Exponential Families," in *Statistics: Applications and New Directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference,* pp. 579–604, edited by J. K. Ghosh and J. Roy. Calcutta: Indian Statistical Institute.

TUKEY, J. W. 1960. "A Survey of Sampling from Contaminated Distributions," in *Contributions to Probability and Statistics,* edited by J. Olkin, W. Hoeffding, S. Ghurye, W. Madow, and H. Mann. Stanford, CA: Stanford University Press.

Discussions on this paper can be submitted until October 1, 2003. The authors reserve the right to reply to any discussion. Please see the Submission Guidelines for Authors on the inside back cover for instructions on the submission of discussions.