© *The Journal of Risk and Insurance*, 2005, Vol. 72, No. 3, 479-495

A Class of Conjugate Priors for Log-Normal Claims Based on Conditional Specification

José María Sarabia
Enrique Castillo
Emilio Gómez-Déniz
Francisco J. Vázquez-Polo

Abstract

In this article, a new methodology for obtaining a premium based on a broad class of conjugate prior distributions, assuming lognormal claims, is presented. The new class of prior distributions arise in a natural way, using the conditional specification technique introduced by Arnold, Castillo, and Sarabia (1998, 1999). The new family of prior distributions is very flexible and contains, as particular cases, many other distributions proposed in the literature. Together with its flexibility, the main advantage of this distribution is that, due to its dependence on a large number of hyperparameters, it allows incorporating a wide amount of prior information. Several methods for hyperparameter elicitation are proposed. Finally, some examples with real and simulated data are given.

Introduction

Actuaries have found (Hewitt and Lefkowitz, 1979; Hogg and Klugman, 1984; among others) that a lognormal distribution is an important model for claim distributions. In many actuarial problems, it is found that although losses do not fit normal curves, mainly because they are skewed toward the upper boundaries, log losses provide a good fit. The lognormal distribution then appears as a natural distribution to be incorporated in credibility theory. Also, the concept of log credibility has been used in several recent papers and applications (see, for example, Landsman and Makov,

José María Sarabia works with Department of Economics, University of Cantabria, Spain; Enrique Castillo works with Department of Applied Mathematics and Computational Sciences, University of Cantabria, Spain; Emilio Gómez-Désniz and Francisco J. Vázquez-Polo work with Department of Quantitative Methods, University of Las Palmas de Gran Canaria, Spain. The authors can be contacted via e-mail: sarabiaj@unican.es; castie@unican.es; egomez@dmc.ulpgc.es; and fjvpolo@dmc.ulpgc.es. The authors want to thank the referees for suggesting clarifications and improvements. They also thank the MCYT (Ministerio de Ciencia y Tecnología, Spain) for partial funding of this research (authors from Cantabria under project BEC2000–1186, and authors from Canary Islands under project BEC2001–3774).

1999; Klugman et al., 1998, chapter 5), where the model assumes that claims follow a lognormal distribution. A careful study of the lognormal distribution is contained in Johnson, Kotz, and Balakrishnan (1994, chapter 14).

Therefore, assume that $\underset{\sim}{\theta}$ is a risk parameter characterizing a member of a risk collective, and that the distribution of his claim X given $\underset{\sim}{\theta}$ is lognormal with probability density function:

$$f(x \mid \underset{\sim}{\theta}) = \frac{\tau^{1/2}}{x\sqrt{2\pi}} \exp\left\{-\frac{\tau}{2}(\log x - \mu)^2\right\} I(x > 0), \tag{1}$$

where $\underset{\sim}{\theta} = (\mu, \tau), \tau = \frac{1}{\sigma^2}$ is the precision parameter, and I is the indicator function. This distribution is denoted by $X \sim \mathcal{LN}(\mu, \tau)$. If $\underset{\sim}{x} = (x_1, \ldots, x_n)$ is a lognormal data sample, the likelihood becomes,

$$f(\underset{\sim}{x}; \mu, \tau) = \frac{\tau^{n/2}}{\left(\prod_{i=1}^{n} x_i\right)(2\pi)^{n/2}} \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}(\log x_i - \mu)^2\right\} I(\underset{\sim}{x} > 0). \tag{2}$$

From a Bayesian framework we have the model,

$$\underset{\sim}{X} \mid \mu, \tau \sim \mathcal{LN}(\mu, \tau), \tag{3}$$

$$(\mu, \tau) \sim \pi(\mu, \tau), \tag{4}$$

and a prior density $\pi(\mu, \tau)$ has to be specified for (μ, τ). If conjugate prior distributions are adopted, the classical solution consists of specifying as prior the normal-gamma distribution, which satisfies (DeGroot, 1970; Jewell, 1974; Padgett and Wei, 1977):

$$\mu \mid \tau \sim \mathcal{N}(m, k\tau^{-1}), \tag{5}$$

$$\tau \sim \mathcal{G}(\alpha, \beta), \tag{6}$$

where $\mathcal{N}(\mu, \sigma^2)$ represents a normal distribution with mean μ and variance σ^2 and $\mathcal{G}(\alpha, \beta)$ represents a classical gamma distribution with probability density function proportional to $x^{\alpha-1}e^{-\beta x}$. This prior distribution has two important properties.

1. It is a conjugate prior of Equation (1).
2. Only four parameters (m, k, α, and β) require elicitation.

The first of these properties establishes a congruent model, presents important computational advantages and, in addition, has a tradition in actuarial practice. Along with this mathematical convenience, the use of conjugate structure functions can lead to unrealistic practical situations. For this reason, it is always advisable to evaluate the model for a class of prior densities in a robustness analysis to explore if these distributions are not exerting undue influence on the overall conclusions (Gómez, Hernández, and Vázquez-Polo, 2000, 2002a; Gómez et al., 2002b).

With respect to the second property, four parameters can be insufficient in practice. Normally the user has a great deal of available information for obtaining the prior distribution $\pi(\underline{\theta})$. On the other hand, the particular parameter structure (i.e., $\mathbb{E}(\mu \mid \tau)$ constant and $\text{Var}(\mu \mid \tau) = k\tau^{-1}$) is a direct consequence of the mathematical constraints for having conjugate distributions.

In this article, a new class of prior distributions that *arises* in a natural way when using the conditional specification technique proposed by Arnold, Castillo, and Sarabia (1998, 1999), is introduced. For an introduction to conditionally specified models see Arnold, Castillo, and Sarabia (2001).

The new family of prior distributions is very flexible and contains, as particular cases, many other distributions proposed in the literature. Together with its flexibility, the main advantage of this distribution is that, due to its dependence on a large number of hyperparameters, it allows incorporating a wide variety of prior information. Several methods for hyperparameter assessment are proposed.

The article is organized as follows. The prior distributions based on conditional specifications and its properties are studied in the following section. The next section presents the Bayesian analysis with the new class of priors, together with several proposals for hyperparameter elicitation. This is followed by some examples (with real and simulated data), which are given in order to prove that the proposed methodology is easily automated, widely applicable and flexible with respect to the choice of the structure function. The following section presents the immediate extension to the sampling model generated by the Pareto distribution. The last section contains discussions as well as conclusions.

PRIOR DISTRIBUTIONS BASED ON CONDITIONAL SPECIFICATION

A more flexible and conjugate prior distribution for the model (3)–(4) is based on the following observation. If τ is a known parameter, the normal distribution is a conjugate distribution for μ, whereas if μ is a known parameter, the gamma distribution is conjugate for τ. Consequently, it has sense to ask for the most general prior distribution $\pi(\mu, \tau)$ such that its conditional distributions $\mu \mid \tau$ be normal and the conditional distributions $\tau \mid \mu$ be gammas, such that

$$\mu \mid \tau \sim \mathcal{N}(m(\tau), s^2(\tau)), \tag{7}$$

$$\tau \mid \mu \sim \mathcal{G}(\alpha(\mu), \beta(\mu)), \tag{8}$$

where $m(\tau) : \mathbb{R}^+ \longrightarrow \mathbb{R}, s^2(\tau) : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$, and $\alpha(\mu), \beta(\mu) : \mathbb{R} \longrightarrow \mathbb{R}^+$ (see Castillo and Galambos, 1987; Castillo, Galambos, and Sarabia, 1990). Then, we have the following theorem.

Theorem 1: *The most general bivariate distribution with conditional distributions (7) and (8) is given by*

$$\pi(\mu, \tau; \Lambda) = \exp\left\{\underline{u}_\mu \Lambda \underline{v}_\tau^T\right\}; \quad \mu \in \mathbb{R}, \quad \tau \in \mathbb{R}^+, \tag{9}$$

where the vectors $\underline{u}_\mu$ *and* $\underline{v}_\tau$ *are given by*

$$\underline{u}_\mu = (1, \mu, \mu^2), \tag{10}$$

$$\underline{v}_\tau = (1, \tau, \log\tau), \tag{11}$$

and $\Lambda = \{\lambda_{ij}\}$, $i, j = 0, 1, 2$. *The parameter* λ_{00} *is the normalizing constant and must be chosen to satisfy* $\iint \pi(\mu, \tau; \Lambda)\, d\mu\, d\tau = 1$. *The parameters* $\{\lambda_{ij}\}$ *must be selected to satisfy* $\int \pi(\mu, \tau; \Lambda)\, d\mu < \infty$ *or* $\int \pi(\mu, \tau; \Lambda)\, d\tau < \infty$.

Proof: Since both conditional distributions belong to the exponential (normal and gamma distributions) family, this is a simple particular case of theorem 4.1 in Arnold, Castillo, and Sarabia (1999). Q.E.D.

We call bivariate normal–gamma conditionals distribution (BCNG) to this prior distribution, denoted as $(\mu, \tau) \sim \mathcal{BCNG}(\Lambda)$, which was proposed by Arnold, Castillo, and Sarabia (1998) in the context of normal likelihoods. The case of lognormal likelihoods is new and has not been considered previously.

Properties of the BCNG Distribution

In this section, the basic properties of the $\mathcal{BCNG}$ distribution are studied. They are derived directly from Equation (9), using the following notation:

$$m_1(\tau) = \lambda_{10} + \lambda_{11}\tau + \lambda_{12}\log\tau, \tag{12}$$

$$m_2(\tau) = \lambda_{20} + \lambda_{21}\tau + \lambda_{22}\log\tau, \tag{13}$$

$$\alpha(\mu) = 1 + \lambda_{02} + \lambda_{12}\mu + \lambda_{22}\mu^2, \tag{14}$$

$$\beta(\mu) = -\big(\lambda_{01} + \lambda_{11}\mu + \lambda_{21}\mu^2\big). \tag{15}$$

Joint Probability Density Function. Expanding Equation (9), the following bivariate density function is obtained

$$\pi(\mu, \tau; \Lambda) = \exp\big\{\lambda_{00} + \lambda_{10}\mu + \lambda_{20}\mu^2 + \lambda_{01}\tau + \lambda_{02}\log\tau + \lambda_{11}\mu\tau + \lambda_{12}\mu\log\tau + \lambda_{21}\mu^2\tau + \lambda_{22}\mu^2\log\tau\big\}, \tag{16}$$

where $\mu \in \mathbb{R}, \tau \in \mathbb{R}^+$, and λ_{00} is a function of the other parameters.

If we are willing to accept improper priors, then conditions on the above parameters are not required. In addition, for Equation (9) to be a genuine density function, several constraints must be satisfied. One can immediately observe that

$$\lambda_{01} < 0, \quad \lambda_{02} > -1, \quad \lambda_{21} < 0, \quad \lambda_{22} > 0, \tag{17}$$

$$\lambda_{12}^2 < 4\lambda_{22}(\lambda_{02} + 1), \tag{18}$$

$$\lambda_{11}^2 < 4\lambda_{21}\lambda_{01}, \tag{19}$$

$$\lambda_{20} + \lambda_{22}\left[\log(-\lambda_{22}/\lambda_{21}) - 1\right] < 0. \tag{20}$$

The normal-gamma conditionals distribution contains, as particular cases, the following submodels.

1. The normal-gamma distribution given in Equations (5) and (6). This conjugate classical situation corresponds to $\lambda_{10} = \lambda_{20} = \lambda_{12} = \lambda_{22} = 0$ and $\lambda_{21} < 0$, $\lambda_{02} > -1$, $\lambda_{01} < 0$, $\lambda_{11}^2 < 4\lambda_{01}\lambda_{21}$.
2. The independence case. This situation corresponds to the case $\lambda_{11} = \lambda_{12} = \lambda_{21} = \lambda_{22} = 0$ and $\lambda_{20} < 0$, $\lambda_{02} > -1$, $\lambda_{01} < 0$.

Conditional and Marginal Distributions. The conditional distributions of Equation (9) are Equations (7) and (8), with means and variances given by

$$\mathbb{E}(\mu \mid \tau) = m(\tau) = \frac{-m_1(\tau)}{2m_2(\tau)}, \tag{21}$$

$$\mathrm{Var}(\mu \mid \tau) = s^2(\tau) = \{-2m_2(\tau)\}^{-1}, \tag{22}$$

$$\mathbb{E}(\tau \mid \mu) = \frac{\alpha(\mu)}{\beta(\mu)}, \tag{23}$$

$$\mathrm{Var}(\tau \mid \mu) = \frac{\alpha(\mu)}{\beta(\mu)^2}. \tag{24}$$

Now, integrating Equation (9), the following expressions for the marginal distributions are obtained

$$\pi_1(\mu; \Lambda) = \exp\left\{\lambda_{00} + \lambda_{10}\mu + \lambda_{20}\mu^2\right\}\frac{\Gamma(\alpha(\mu))}{[\beta(\mu)]^{\alpha(\mu)}}, \tag{25}$$

$$\pi_2(\tau; \Lambda) = \exp\{\lambda_{00} + \lambda_{01}\tau + \lambda_{02}\log\tau\}\exp\left\{\frac{-m_1(\tau)^2}{4m_2(\tau)}\right\}\sqrt{\frac{\pi}{-m_2(\tau)}}. \tag{26}$$

showing that, although the conditional distributions $\mu \mid \tau$ and $\tau \mid \mu$ are normal and gamma, respectively, the marginal distributions of μ and τ do not follow in general these two types of models. Note that Equations (25) and (26) are normal and gamma, respectively, only when $\lambda_{11} = \lambda_{12} = \lambda_{21} = \lambda_{22} = 0$, that is, in the independence case.

Also, the mode is the solution of the system of equations

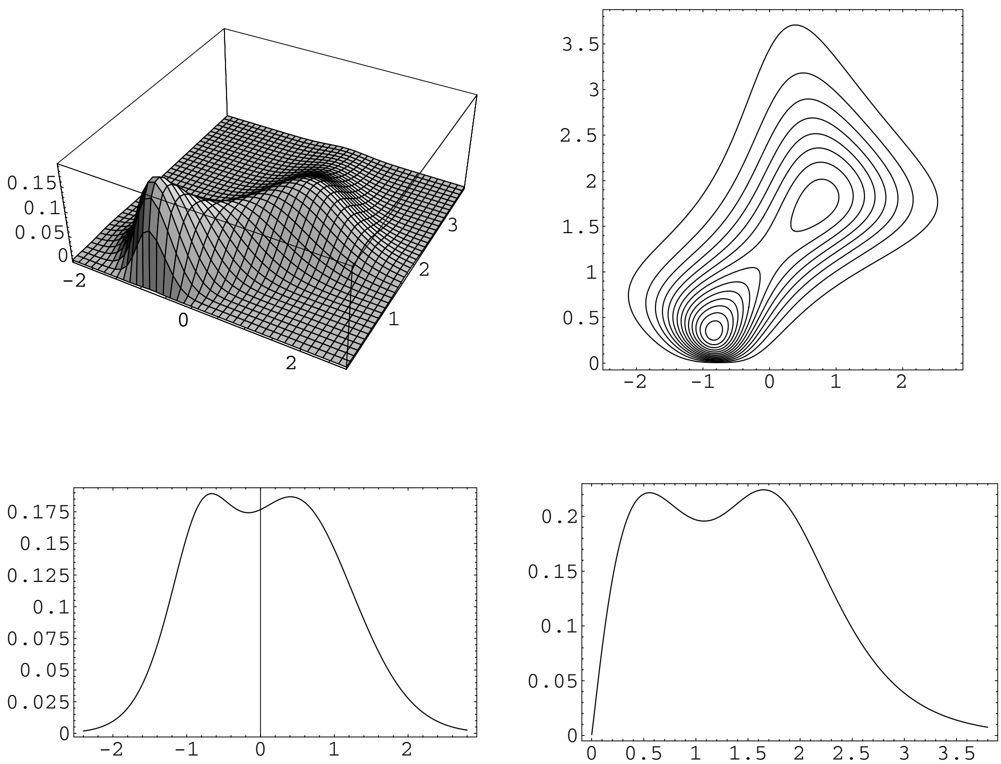
$$\lambda_{10} + 2\lambda_{20}\mu + \lambda_{11}\tau + \lambda_{12}\log\tau + 2\lambda_{21}\mu\tau + 2\lambda_{22}\mu\log\tau = 0, \tag{27}$$

$$\lambda_{02} + \lambda_{12}\mu + \lambda_{22}\mu^2 + \tau\left(\lambda_{01} + \lambda_{11}\mu + \lambda_{21}\mu^2\right) = 0. \tag{28}$$

Replacing τ from Equation (28) in Equation (27) one obtains one equation in μ. Numerical computations show that the solution of Equations (27) and (28) need not be

FIGURE 1

A Bivariate Conditionals Normal Gamma Distribution With Two Modes. The pdf of an $\mathcal{BCNG}$ Distribution With $\lambda_{10} = 0.7$, $\lambda_{20} = 1.4$, $\lambda_{01} = -2.35$, $\lambda_{02} = 2.7$, $\lambda_{11} = -1.3$, $\lambda_{12} = 4.8$, $\lambda_{21} = -2.3$, and $\lambda_{22} = 3.3$ (Upper Left) and the Corresponding Contour Plots (Upper Right). The pdf of the μ Marginal (Lower Left) and the τ Marginal (Lower Right). All the Figures Correspond to the pdf's Divided by $\exp(\lambda_{00})$



unique, and can lead to bimodality. One example of this situation is illustrated in Figure 1, which shows a bivariate normal-gamma conditionals distribution with two modes. Multimodality appears in other models with conditional specification (see, for example, Arnold et al., 2000; Arnold, Castillo, and Sarabia, 2001).

Dependence. The classical normal–gamma model given in Equations (5) and (6) admits only negative nonlinear correlations. In order to prove that $\rho(\mu, \tau) \leq 0$, it suffices to observe that

$$\Pr(\mu > \mu_0 \mid \tau = \tau_0) = 1 - \Phi\left(\sqrt{\frac{\tau_0}{k}}(\mu_0 - m)\right),$$

is decreasing in τ_0 for every μ_0. Now, using theorem 4.2 of Barlow and Proschan (1975), it follows that μ and τ are negative-quadrant dependent, which implies a negative correlation.

However, the BCNG distributions admit positive and negative correlations because

$$\Pr(\mu > \mu_0 \mid \tau = \tau_0) = 1 - \Phi\left(\mu_0\sqrt{-2m_2(\tau_0)} - \frac{m_1(\tau_0)}{\sqrt{-2m_2(\tau_0)}}\right),$$

can be increasing or decreasing in τ_0.

BAYESIAN ANALYSIS WITH THE BCNG PRIOR

In order to obtain a posterior distribution, the likelihood in Equation (2) can be rewritten as

$$f(\underset{\sim}{x} \mid \underset{\sim}{\theta}) = (2\pi)^{-n/2} \prod_{i=1}^{n} \left(x_i^{-1}\right) \exp\left\{\frac{n}{2}\log\tau - \frac{\sum_{i=1}^{n}(\log x_i)^2}{2}\tau - \frac{n}{2}\mu^2\tau + \mu\tau\sum_{i=1}^{n}\log x_i\right\}, \tag{29}$$

then if a prior

$$(\mu, \tau) \sim \mathcal{BCNG}\left(\Lambda^{(0)}\right), \tag{30}$$

is chosen and combined with Equation (29), one obtains the following posterior distribution

$$(\mu, \tau) \mid \underset{\sim}{X} \sim \mathcal{BCNG}(\Lambda^*), \tag{31}$$

where the hyperparameter vector $\Lambda^{(0)}$ is updated to Λ^* using the expressions in Table 1. Note that only four of the eight parameters are modified by data—λ_{01}, λ_{02},

TABLE 1
Hyperparameter Updating of the Prior Family (16), Combined With Likelihood (29)

Parameter	Prior Value	Posterior Value
λ_{10}	$\lambda_{10}^{(0)}$	$\lambda_{10}^{\star} = \lambda_{10}^{(0)}$
λ_{20}	$\lambda_{20}^{(0)}$	$\lambda_{20}^{\star} = \lambda_{20}^{(0)}$
λ_{01}	$\lambda_{01}^{(0)}$	$\lambda_{01}^{\star} = \lambda_{01}^{(0)} - \frac{1}{2}\sum_{i=1}^{n}\log^2 x_i$
λ_{02}	$\lambda_{02}^{(0)}$	$\lambda_{02}^{\star} = \lambda_{02}^{(0)} + \frac{n}{2}$
λ_{11}	$\lambda_{11}^{(0)}$	$\lambda_{11}^{\star} = \lambda_{11}^{(0)} + \sum_{i=1}^{n}\log x_i$
λ_{12}	$\lambda_{12}^{(0)}$	$\lambda_{12}^{\star} = \lambda_{12}^{(0)}$
λ_{21}	$\lambda_{21}^{(0)}$	$\lambda_{21}^{\star} = \lambda_{21}^{(0)} - \frac{n}{2}$
λ_{22}	$\lambda_{22}^{(0)}$	$\lambda_{22}^{\star} = \lambda_{22}^{(0)}$

λ_{11}, and λ_{21}. The remaining parameters do not change, but their contribution to the posterior would be eventually swamped by large data sets. The presence of these parameters allows a larger flexibility when selecting the prior distribution. This flexibility is not present in the classical model (5)–(6).

Parameter Estimation

In the conditional context, Gibbs sampling is a natural estimation mode. Assume that one is interested in approximating the posterior moments of a given function of μ and τ, say $\delta(\mu, \tau)$. Then, to approximate $\mathbb{E}(\delta(\mu, \tau) \mid \underset{\sim}{x})$ we generate random values of $\mu_1, \tau_1, \mu_2, \tau_2, \ldots, \mu_{m_0+m}, \tau_{m_0+m}$ using the conditional distributions

$$\mu \mid (\tau, \underset{\sim}{x}) \sim \mathcal{N}(m^*(\tau), s^{2*}(\tau)), \tag{32}$$

$$\tau \mid (\mu, \underset{\sim}{x}) \sim \mathcal{G}(\alpha^*(\mu), \lambda^*(\mu)), \tag{33}$$

where the expressions for obtaining the λ_{ij}^* are given in Table 1. Finally, if we allow for m_0 observations for the sampler to stabilize, before starting the actual sampling, we have the estimator

$$\mathbb{E}(\delta(\mu, \tau) \mid \underset{\sim}{x}) \approx \frac{1}{m} \sum_{i=m_0+1}^{m_0+m} \delta(\mu_i, \tau_i). \tag{34}$$

Other estimators are also possible. The posterior distribution is a member of an exponential family, so numerical determination of these posterior expectations is possible. Alternatively, we can use the mode of the posterior distributions, solving the system (27)–(28).

Hyperparameter Elicitation

The selected prior distribution has been built based on the conditional distributions of μ given τ and τ given μ. Consequently, it seems reasonable to perform the hyperparameter elicitation based on conditional information. Some of these procedures have been proposed by Arnold, Castillo, and Sarabia (1998) with normal and Pareto likelihoods. We assume that several items of information could be available, such as:

- Information about the percentiles of $\mu \mid \tau$ and $\tau \mid \mu$. Assume that a total of n_1 percentiles $x_i^{(1)}$ of the conditional distribution $\mu \mid \tau$, and n_2 percentiles $x_j^{(2)}$ of $\tau \mid \mu$ are available, i.e.,

$$F_{\mu|\tau}\left(x_i^{(1)}\right) = p_i^{(1)}, \quad i = 1, 2, \ldots, n_1 \tag{35}$$

$$F_{\tau|\mu}\left(x_j^{(2)}\right) = p_j^{(2)}, \quad j = 1, 2, \ldots, n_2 \tag{36}$$

where $0 \le p_i^{(1)}, p_j^{(2)} \le 1$. Then, Equations (35) and (36) are solved for the parameters $\{\lambda_{ij}\}$. Note that the Equations (35) and (36) are nonlinear in λ_{ij} and numerical methods are required.

- Information on conditional moments. Assume that we have information about the mean and variance of the two conditional distributions $\mu \mid \tau$ and $\tau \mid \mu$, i.e.,

$$\mathbb{E}(\mu \mid \tau_i) = a_i, \quad i = 1, 2, \ldots, n_1 \tag{37}$$

$$\mathrm{Var}(\mu \mid \tau_i) = b_i, \quad i = 1, 2, \ldots, n_1 \tag{38}$$

$$\mathbb{E}(\tau \mid \mu_i) = c_i, \quad i = 1, 2, \ldots, n_2 \tag{39}$$

$$\mathrm{Var}(\tau \mid \mu_i) = d_i, \quad i = 1, 2, \ldots, n_2 \tag{40}$$

where the values of a_i, b_i, c_i, and d_i are known. Then, replacing Equations (21) and (24) in Equations (37)–(40), we get

$$\lambda_{10} + \lambda_{11}\tau_i + \lambda_{12}\log\tau_i = \frac{a_i}{b_i}, \tag{41}$$

$$\lambda_{20} + \lambda_{21}\tau_i + \lambda_{22}\log\tau_i = -\frac{1}{2b_i}, \tag{42}$$

$$\lambda_{01} + \lambda_{11}\mu_j + \lambda_{21}\mu_j^2 = -\frac{c_j}{d_j}, \tag{43}$$

$$\lambda_{02} + \lambda_{12}\mu_j + \lambda_{22}\mu_j^2 = \frac{c_j^2}{d_j} - 1, \tag{44}$$

that is, a linear system in $\underset{\sim}{\lambda} = (\lambda_{01}, \ldots, \lambda_{22})$, which can be written in the form $\underset{\sim}{\lambda} M = A$. Then, least squares estimates of the parameters subject to constraints (17)–(20) are then obtainable by using regression techniques. In next section, an example of this elicitation method is provided.
- Combined information about percentiles and moments of $\mu \mid \tau$ and $\tau \mid \mu$. This method combines information of the type (35)–(36) with (41)–(44). Thus, nonlinear systems are again obtained.
- Noninformative priors. If the expert expresses inability to provide any conditional percentiles or moments of this prior, it would be appropriate to use a uniform prior for (μ, τ). A possibility is to use the Jeffreys prior distribution (see "Example 1").

NUMERICAL EXPERIMENTS

In this section, we describe the implementation of the model incorporating real and simulated data. These examples illustrate that the proposed methodology is easily automated, widely applicable, and flexible with respect to the choice of the function structure. All computations and simulations were done using WinBUGS (Spiegelhalter, Thomas, and Best, 1999). Appendix contains the WinBUGS code for one of the models developed here. We used three parallel chains and a single long chain for diagnostic assessment (checked using CODA software). A total of 40,000 iterations were carried out (after a burn-in period), taking only a few seconds on a 400 MHz Pentium personal computer. Note that the gamma (0.001, 0.001) prior distribution is a popular choice for a diffuse prior information. Thus, near to improper priors or noninformative priors are also supported in WinBUGS and constant λ_{00} is not required to ensure convergence.

TABLE 2
Losses (in Millions) From Hurricanes in the Period 1949–1980 (Hogg and Klugman, 1984)

6.766	7.123	10.562	14.474	15.351	16.983	18.383
19.030	25.304	29.112	30.146	33.727	40.596	41.409
47.905	49.397	52.600	59.917	63.123	77.809	102.942
103.217	123.680	140.136	192.013	198.446	227.338	329.511
361.200	421.680	513.586	545.779	750.389	863.881	1638.000

Example 1

In this first example, we take data from Hogg and Klugman (1984) related to the data compiled by the American Insurance Association from 1949 through 1980, corresponding to the total damage done by a catastrophe process (hurricane in this situation). The amounts, in adjusted dollars, are presented in Table 2.

In order to select one model among the fitted distributions, several loss models have been proposed in Hogg and Klugman (1984), being the lognormal model with the highest value of $-\ln L$. Therefore the lognormal model plausibly fits the data. Maximum likelihood estimates of parameters are $\hat{\mu} = 17.953$ and $\hat{\sigma} = 1.6028$. Two possible cases of expert knowledge are considered:

1. Noninformative priors: For comparative purposes with maximum likelihood estimation, a (improper) vague prior has been chosen. Additionally, if little is known *a priori* about the quantities of interest, several methods have been developed for specification of diffuse priors (Bernardo and Smith, 1994; Jeffreys, 1967; Gelman et al., 1995; among others). The Jeffreys prior distribution used here is given by

$$\pi^J(\mu, \tau) \propto |I(\mu, \tau)|^{1/2} \propto \tau^{-1/2} = \exp\left\{-\frac{1}{2}\log\tau\right\}, \tag{45}$$

   where $I(\mu, \tau)$ is the Fisher information matrix, that is, minus the expectation of the matrix of second-order partial derivatives of $\log f(x \mid \mu, \tau)$. Together with this prior we developed the example by using other alternative diffuse prior that also produce proper posteriors

$$\pi^{NI}(\mu, \tau) \propto \frac{1}{\tau} \propto \exp\{-\log\tau\}. \tag{46}$$

2. Prior conditional moments: With the purpose of showing how the elicitation techniques work, the following experiment has been done. For fixed values of $\tau \in \{1, 2\}$ a normal sample with mean $\mu = 14$ of size 100 has been simulated. According to Equations (37) and (38), the sample mean and variance give the values of a_1 and b_1, respectively. The resulting values were $a_1 = 14.12$ and $b_1 = 1.10$, for $\tau = 1$, and $a_2 = 14.04$ y $b_2 = 1.92$, for $\tau = 2$. Conversely, for fixed values of $\mu \in \{10, 15\}$, 100 samples of a gamma distribution were simulated for $\mathcal{G}(4, 0.5)$ and $\mathcal{G}(1, 1)$,

TABLE 3
MCMC Parameter Estimation From the Hurricane Data Set

	Parameter					
	μ		τ		$e^{\mu+\tau^{-1}/2} \times 10^6$	
Prior Density	Post. Mean	95% Bayesian Interval	Post. Mean	95% Bayesian Interval	Post. Mean	95% Bayesian Interval
$\pi^J(\mu, \tau)$	18.15	(17.68, 18.63)	0.5187	(0.3092, 0.7824)	225.6	(117.5, 459.3)
$\pi^{NI}(\mu, \tau)$	18.15	(17.68, 18.62)	0.5328	(0.32, 0.8001)	218.4	(115.9, 435.0)
$\pi^I(\mu, \tau)$	17.86	(17.24, 18.45)	0.2966	(0.1544, 0.4837)	484.6	(135.0, 1704.0)

which allowed obtaining the sample mean and variances as approximations to c_i and d_i's, as given in Equations (39) and (40). The resulting values were $c_1 = 1.90$, $d_1 = 0.9347$, $c_2 = 1.13$, and $d_2 = 1.39$. Then, solving the system of linear Equations (41)–(44), we obtained $\lambda_{01} = 3.182$, $\lambda_{02} = 9.36$, $\lambda_{10} = 18.64$, $\lambda_{11} = -5.81$, $\lambda_{12} = 0.41$, $\lambda_{20} = -0.70$, $\lambda_{21} = 0.24$, $\lambda_{22} = -0.07$. Then, the prior density, $\pi^I(\mu, \tau)$ was obtained by substitution of these values into Equation (16).

Table 3 shows the obtained results.

The third column corresponds to the posterior distribution of the fair premium, under net premium principle and lognormal claims data, $\delta(\mu, \tau) = \mathbb{E}(X \mid \mu, \tau)$, given n individual experience $x_1, x_2, \ldots, x_n$. Roughly speaking, this is the appropriate premium to cover next risk X_{n+1}. Although there is little difference in the three priors considered with respect to point estimates for μ and τ, the largest differences appear in the 95 percent central posterior Bayesian intervals for fair premium. The expert's opinion, given in $\pi^I(\mu, \tau)$, makes τ more precise and reduce the width of the confidence intervals.

Example 2

The following simulation exercise was developed for inference purposes. We have simulated data from $\mathcal{LN}(\mu, \tau)$ with $\mu = 3$, and $\tau = 1$. For illustrations, several prior situations are presented:

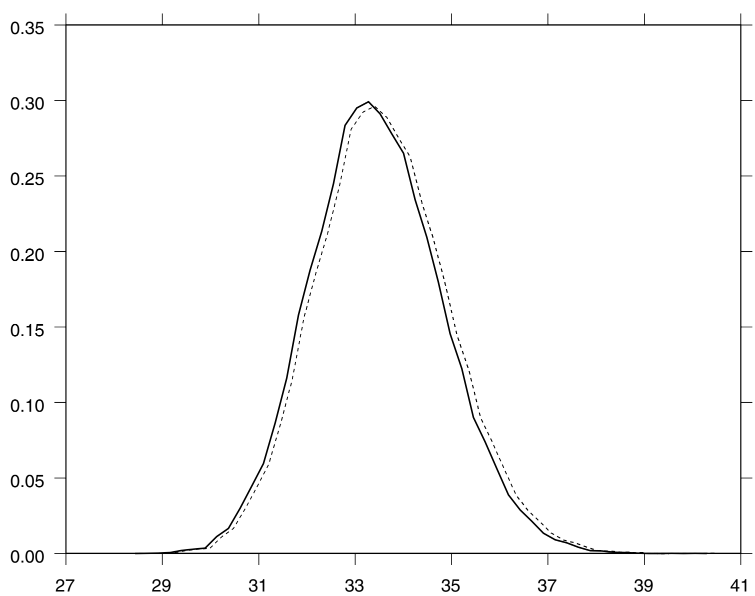
- Normal-gamma prior: This particular situation corresponds to the following prior parameter elicitation $\lambda_{10} = \lambda_{20} = \lambda_{12} = \lambda_{22} = 0, \lambda_{02} = 2, \lambda_{01} = -1, \lambda_{11} = \frac{1}{2}, \lambda_{21} = -\frac{1}{2}$.
- Independence: $\lambda_{11} = \lambda_{12} = \lambda_{21} = \lambda_{22} = 0, \lambda_{02} = 1, \lambda_{01} = -1, \lambda_{10} = \frac{1}{2}, \lambda_{20} = -\frac{1}{2}$.
- Near independence: We considered this situation by making nonnull the parameter λ_{11} involved in the cross-product between μ and τ.

Table 4 presents all posterior summaries for the parameters of interest. In every case, the reported posterior mean and central Bayesian posterior intervals were quite similar to those given by the classical normal-gamma model. Hence, this limited sensitivity analysis suggests a reasonable robustness. Figure 2 illustrates this small sensitivity to two assumptions.

TABLE 4
Results From Simulated Data

$E\{\delta(\mu,\tau)\mid\text{data}\}$	95% CI	$E\{\mu\mid\text{data}\}$	95% CI	$E\{\tau\mid\text{data}\}$	95% CI
		Normal-gamma scenario			
33.32	(30.85, 36.03)	2.984	(2.92, 3.047)	0.9605	(0.8784, 1.048)
		Independence scenario			
33.45	(30.95, 36.19)	2.984	(2.92, 3.047)	0.9532	(0.8616, 1.04)
		"Nearindependence" scenario ($\lambda_{11} = \varepsilon \to 0$)			
		$\varepsilon = 10^{-7}$			
33.44	(30.92, 36.17)	2.983	(2.92, 3.047)	0.9534	(0.8719, 1.039)
		$\varepsilon = 10^{-6}$			
33.45	(30.93, 36.23)	2.984	(2.92, 3.04)	0.9532	(0.8719, 1.039)
		$\varepsilon = 10^{-5}$			
33.45	(30.96, 36.22)	2.984	(2.919, 3.048)	0.9532	(0.8727, 1.038)
		$\varepsilon = 10^{-4}$			
33.44	(30.92, 36.17)	2.983	(2.92, 3.047)	0.9534	(0.8719, 1.039)
		$\varepsilon = 10^{-3}$			
33.44	(30.92, 36.17)	2.983	(2.92, 3.047)	0.9534	(0.8719, 1.039)
		$\varepsilon = 10^{-2}$			
33.45	(30.95, 36.19)	2.984	(2.92, 3.047)	0.9532	(0.8716, 1.04)
		$\varepsilon = 10^{-1}$			
33.46	(30.96, 36.2)	2.984	(2.92, 3.047)	0.9527	(0.8711, 1.039)
		$\varepsilon = 10^{0}$			
33.58	(31.06, 36.35)	2.985	(2.921, 3.048)	0.9478	(0.8667, 1.034)

FIGURE 2
Posterior Density of $\delta(\mu,\tau) = \exp(\mu + \frac{1}{2}\tau^{-1})$ in Two Limit Situations: Independence and Dependence ($\lambda_{11}^{(0)} = 1$) (dashed line) Settings



EXTENSIONS: THE PARETO DISTRIBUTION

The model presented here can be easily extended to other sample distributions showing lack of symmetry, as the lognormal distribution, and with applications to actuarial science, such as the Pareto distribution. This is our first proposal for extension. This case was considered by Arnold, Castillo, and Sarabia (1998). Then, consider data from a classical Pareto distribution with shape parameter α, precision parameter τ, and probability density function

$$f(x;\alpha,\tau) = \begin{cases} \alpha\tau(\tau x)^{-(\alpha+1)}, & \tau x > 1, \\ 0, & \tau x < 1. \end{cases} \tag{47}$$

The natural parameter space is $\alpha > 0$, $\tau > 0$. If X has a density (47) we write $X \sim \mathcal{P}(\alpha,\tau)$. Now, if α were known, a natural conjugate prior density for τ would be a Pareto density, and if τ were known, then the natural conjugate prior family would be the gamma distribution. Consequently, we ask for the most general bidimensional distribution of (α, τ) such that the conditional distributions $\tau \mid \alpha$ and $\alpha \mid \tau$ be Pareto and gamma distributions, respectively. This prior distribution is given by

$$\pi(\alpha,\tau;\Lambda) = \exp\left\{(1,\alpha,\log\alpha)\Lambda(1,\log\tau)^T\right\} I(\alpha > 0, \tau\lambda_{30} > 1), \tag{48}$$

where $\Lambda = \{\lambda_{ij}\}$, $i = 0, 1, 2$; $j = 0, 1$ is a 3×2 parameter matrix, and λ_{00} is the normalizing constant. We call bivariate gamma Pareto conditionals distribution (BCGP) to this distribution, which is denoted $(\alpha,\tau) \sim \mathcal{BCGP}(\Lambda)$. The conditional distributions are

$$\tau \mid \alpha \sim \mathcal{P}(a_0(\alpha), b_0(\alpha)), \tag{49}$$

$$\alpha \mid \tau \sim \mathcal{G}(c_0(\tau), d_0(\tau)), \tag{50}$$

where

$$a_0(\alpha) = -(1 + \lambda_{01} + \lambda_{11}\alpha + \lambda_{21}\log\alpha), \tag{51}$$

$$b_0(\alpha) = \lambda_{30}, \tag{52}$$

$$c_0(\tau) = 1 + \lambda_{20} + \lambda_{21}\log\tau, \tag{53}$$

$$d_0(\tau) = -(\lambda_{10} + \lambda_{11}\log\tau). \tag{54}$$

Integrating in Equation (48) we obtain the marginal distributions

$$\pi_1(\alpha;\Lambda) = \exp\{\lambda_{00} + \lambda_{10}\alpha + \lambda_{20}\log\alpha\}\frac{\lambda_{20}^{a_0(\alpha)}}{a_0(\alpha)} I(\alpha > 0), \tag{55}$$

$$\pi_2(\tau;\Lambda) = \exp\{\lambda_{00} + \lambda_{01}\log\tau\}\frac{\Gamma(c_0(\tau))}{d_0(\tau)^{c_0(\tau)}} I(\tau\lambda_{30} > 1). \tag{56}$$

In order Equation (48) to be a genuine joint pdf, the next constraints must be satisfied

$$\lambda_{21} > 0, \quad \lambda_{11} < 0, \quad \lambda_{30} > 0, \tag{57}$$

$$\lambda_{10} < \lambda_{11} \log(\lambda_{30}), \tag{58}$$

$$\lambda_{20} > \lambda_{21} \log(\lambda_{30}) - 1, \tag{59}$$

$$\lambda_{21}[1 - \log(-\lambda_{21}/\lambda_{11})] > \lambda_{01}. \tag{60}$$

Note that Equations (55) and (56) are not, in general, gamma and Pareto distributions, respectively. In the case $\lambda_{11} = \lambda_{21} = 0$, we obtain the independence case, with gamma and Pareto marginals. If we take a random sample of size n from Equation (47) we obtain the likelihood

$$f(\underset{\sim}{x} \mid \alpha, \tau) = \alpha^n \tau^{-n\alpha} \left(\prod_{i=1}^{n} x_i \right)^{-(\alpha+1)} I(\tau x_{1:n} > 1), \tag{61}$$

where $x_{1:n} = \min\{x_1, \ldots, x_n\}$ represents the minimum of the sample. Now, if we choose

$$(\alpha, \tau) \sim \mathcal{BCGP}(\Lambda), \tag{62}$$

and combine with Equation (61), we obtain the posterior distribution

$$(\alpha, \tau) \mid \underset{\sim}{X} \sim \mathcal{BCGP}(\Lambda^*) \tag{63}$$

where the hyperparameter vector Λ^0 is updated to $\Lambda^\star$, according to Table 5. Elicitation of hyperparameters can be done using techniques similar to those in section on "Hyperparameter Elicitation" (see Arnold, Castillo, and Sarabia, 1998). Finally, if we are interested in the estimation of a function of $\delta(\alpha, \tau)$, we use an estimator similar to Equation (34).

TABLE 5
Hyperparameter Updating of the Prior Family (62), Combined With Likelihood (61)

Parameter	Prior Value	Posterior Value
λ_{10}	$\lambda_{10}^{(0)}$	$\lambda_{10}^\star = \lambda_{10}^{(0)} - \sum_{i=1}^n \log x_i$
λ_{20}	$\lambda_{20}^{(0)}$	$\lambda_{20}^\star = \lambda_{20}^{(0)} + n$
λ_{01}	$\lambda_{01}^{(0)}$	$\lambda_{01}^\star = \lambda_{01}^{(0)}$
λ_{11}	$\lambda_{11}^{(0)}$	$\lambda_{11}^\star = \lambda_{11}^{(0)} - n$
λ_{21}	$\lambda_{21}^{(0)}$	$\lambda_{21}^\star = \lambda_{21}^{(0)}$
λ_{30}	$\lambda_{30}^{(0)}$	$\lambda_{30}^\star = \min\{x_{1:n}, \lambda_{30}^{(0)}\}$

CONCLUSION

It is useful in statistical modelling to have tractable multivariate distributions with given marginals in order to be able to quantify the effect of dependence of the variables in the model. In this article, we have presented this possibility using given conditional distributions. We have focused our experiments on independence structure based on λ_{11}. We have assumed this particular way to move away from the prior independence. Although this is not the only way, the technical approach is feasible, and if absence of robustness can thus be obtained, we can conclude absence of prior independence.

Our article has provided a significant extension to the usual normal-gamma model widely used in the actuarial literature under lognormal sampling scheme. This extension can be of interest from two points of view. First, due to its "conditioned–conjugancy" property, it is very easy to incorporate it into a Markov Chain Monte Carlo routine jointly with the flexibility of the inference method. Second, because the methodology offers several ways of incorporating prior knowledge, and introduces a heuristic procedure for checking independence structures between the parameters of interest.

One important part of the future research will consists of finding credibility formulas (exact or approximate) for this new class of prior distributions. These formulas will help to understand the meaning of the eight parameters appearing in the distribution (9).

APPENDIX

WinBUGS Template

```
model{ for(i in 1:N){

logx[i]<-log(x[i])

logx[i]~dnorm(mu,tau)

}

mu~dnorm(m.mu,tau.mu) tau~dgamma(alpha.tau,lambda.tau)

m1.mu<-lambda10 + lambda11*tau + lambda12*log(tau)

m2.mu<-lambda20+lambda21*tau+lambda22*log(tau)

m.mu<--m1.mu/(2*m2.mu)

sigma2.mu<-1/(-2*m2.mu)

tau.mu<-1/sigma2.mu

alpha.tau<-1+lambda02+lambda12*mu+lambda22*mu*mu

lambda.tau<-lambda01+lambda11*mu+lambda21*mu*mu

media<-exp(mu+(1/tau)/2)

}
```

```
# sample size
list(N=35)

# input lambda's
list(lambda.......)

# starting values
list(mu=1,tau=1)
x[]
6.766
7.123
  .
(31 more data)
  .
863.881
1638.000
END
```

REFERENCES

Arnold, B. C., E. Castillo, and J. M. Sarabia, 1998, Bayesian Analysis for Classical Distributions Using Conditionally Specified Priors, *Sankhya, Series B*, 60: 228-245.

Arnold, B. C., E. Castillo, and J. M. Sarabia, 1999, *Conditional Specification of Statistical Models*, Springer Series in Statistics (New York: Springer Verlag).

Arnold, B. C., E. Castillo, and J. M. Sarabia, 2001, Conditionally Specified Distributions: An Introduction (with discussion), *Statistical Science*, 16(3): 249-274.

Arnold, B. C., E. Castillo, J. M. Sarabia, and L. González-Vega, 2000, Multiple Modes in Densities with Normal Conditionals, *Statistics and Probability Letters*, 49: 355-363.

Barlow, R. E., and F. Proschan, 1975, *Statistical Theory of Reliability and Life Testing: Probability Models* (New York: Holt, Rinehart and Winston).

Bernardo, J. M., and A. F. M. Smith, 1994, *Bayesian Theory* (New York: John Wiley).

Castillo, E., and J. Galambos, 1987, Bivariate Distributions With Normal Conditionals, in *Proceedings of the IASTED International Symposium: Simulation, Modelling and Development—SMD '87*, pp. 59-62, Cairo. ACTA Press.

Castillo, E., J. Galambos, and J. M. Sarabia, 1990, Caracterización de modelos bivariantes con distribuciones condicionadas tipo gamma, *Estadística Española*, 32(124): 439-450.

DeGroot, M. H., 1970, *Optimal Statistical Decisions* (New York: McGraw-Hill).

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, 1995, *Bayesian Data Analysis* (London: Chapman and Hall).

Gómez, E., A. Hernández, and F. J. Vázquez-Polo, 2000, Robust Bayesian Premium Principle in Actuarial Science, *Journal of the Royal Statistical Society, Series D*, 49(2): 241-252.

Gómez, E., A. Hernández, and F. J. Vázquez-Polo, 2002a, Bounds for Ratios of Posterior Expectations: Applications in the Collective Risk Model, *Scandinavian Actuarial Journal*, 1: 37-44.

Gómez, E., A. Hernández, J. M. Pérez, and F. J. Vázquez-Polo, 2002b, Measuring Sensitivity in a Bonus–Malus System, *Insurance: Mathematics and Economics*, 31: 105-113.

Hewitt, C. C. Jr., and B. Lefkowitz, 1979, Methods for Fitting Distributions to Insurance Loss Data, *Proceedings of the Casualty Actuarial Society*, LXVI(125-126): 139-160.

Hogg, R. V., and S. A. Klugman, 1984, *Loss Distributions* (New York: John Wiley & Sons).

Jeffreys, H., 1967, *Theory of Probability* (Oxford University Press).

Jewell, W. S., 1974, Credible Means Are Exact Bayesian for Exponential Families, *Astin Bulletin*, 8: 77-90.

Johnson, N. L., S. Kotz, and N. Balakrishnan, 1994, *Continuous Univariate Distributions*, Volume 1, 2nd edition (New York: John Wiley).

Klugman, S. A., J. J. Panjer, and G. E. Willmot, 1998, *Loss Models. From Data to Decisions*. (New York: Wiley).

Landsman, Z., and U. Makov, 1999, Sequential Credibility Evaluation for Symmetric Location Claim Distributions, *Insurance: Mathematics and Economics*, 24: 291-300.

Padgett, W. J., and L. J. Wei, 1977, Bayes Estimation of Reliability for the Two-Parameter Lognormal Distribution, *Communications in Statistics—Theory and Methods*, 6: 443-457.

Spiegelhalter, D. J., A. Thomas, and N. Best, 1999, *WinBUGS, version 1.2*, MRC Biostatistics Unit: Cambridge, U.K.