risks

MDPI

Article

Mixed Size-Biased Log-Normal Distribution with Truncated Normal Prior and Its Application in Insurance Ratemaking

Taehan Bae [1,*], Jieun Kim [2] and Jae Youn Ahn [2]

[1] Department of Mathematics and Statistics, University of Regina, Regina, SK S4S 0A2, Canada
[2] Department of Statistics, Ewha Womans University, Seoul 03760, Republic of Korea; rlawldms9@ewha.ac.kr (J.K.)
* Correspondence: taehan.bae@uregina.ca

Abstract

In the insurance literature, accurately predicting extreme losses has been a persistent and important problem. Recently, under the modelling framework of weighted distributions, several finite-mixture size-biased distributions, including size-biased Weibull and size-biased truncated log-normal distributions, have gained popularity for modelling heavy-tailed insurance claim data. In this study, unlike existing models, we explicitly account for the individual heterogeneity commonly observed in insurance claims by treating the order of size-biased weighting as a continuous latent variable, thereby constructing a mixed size-biased distribution. In particular, we study the various distributional properties of the mixed log-normal distribution with a truncated normal prior, which serves as a conjugate prior for the size-biased log-normal model. For applications in non-life insurance, we discuss the Bayesian credibility premium and present an estimation of a regression model via the EM algorithm. We further conduct a real-data analysis using insurance loss data, comparing goodness-of-fit and tail risk measures with those of standard heavy-tailed distributions.

Keywords: size-biased distribution; log-normal distribution; truncated normal distribution; Bayesian premium; EM algorithm

1. Introduction

In finance and insurance, precise loss forecasting and modeling are essential. Large losses occur infrequently, but failure to predict them properly can result in severe financial consequences. Therefore, developing models that can effectively capture and predict large losses is a critical task in risk management. Extreme Value Theory (EVT) and heavy-tailed data modelling are both crucial for understanding and managing risk in insurance, particularly when dealing with large and infrequent losses. While both approaches are closely related, EVT aims to accurately quantify the probability of rare and severe events such as the maximum/minimum or losses above a threshold (Embrechts et al. 1997). On the other hand, heavy-tailed data modelling focuses on modelling the entire distribution of losses, with a higher probability of large losses than light-tailed (e.g., exponential) distributions. For the purpose of actuarial ratemaking and reserving, various classes of heavy-tailed distributions are used in the literature (Klugman et al. 2019). Amongst the many heavy-tailed distribution classes, the subexponential class of distributions is most widely used in practice due to its ability to explain the dominant role of a single large loss in aggregate losses (Foss et al. 2011). The subexponential distribution class

Check for updates

Academic Editor: Liqun Diao

Received: 3 February 2026
Revised: 13 March 2026
Accepted: 18 March 2026
Published: 23 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

contains well-known heavy-tailed distributions such as Pareto, Burr, Weibull, and log-normal distributions.

In particular, the log-normal distribution has a long history in various fields, including economics, finance, and engineering, and has been recognized as a valuable model for insurance losses (Robert and Klugman 1984). Its positive skewness and ability to model data with a long right tail make it suitable for representing claim amounts that are typically non-negative and may include occasional large losses. However, the standard log-normal distribution may not always adequately capture the extreme tail behaviour observed in real-world insurance data.

To enhance the flexibility and tail behaviour of the log-normal distribution, researchers have explored various modifications and extensions, including the introduction of size-biased weighting. Size-biased distributions, a class of weighted distributions, have gained popularity in various fields, particularly in ecological studies and reliability analysis, accounting for the preferential sampling of larger or more prominent individuals or events (Gupta and Kirmani 1990; Patil and Rao 1978).

Several recent studies have examined the application of size-biased distributions and mixture models in insurance. In the form of a finite mixture of size-biased exponential distributions, the Erlang mixture model (Lee and Lin 2010; Reynkens et al. 2017; Verbelen et al. 2015; Yin and Lin 2016) is one of the most widely used distribution families for modelling complex insurance losses. This model belongs to a dense class of distributions capable of approximating nearly any positive continuous distribution (Willmot and Lin 2011). However, since the Erlang mixture inherently exhibits exponentially decaying tails, it may not adequately capture extreme losses (Reynkens et al. 2017).

To address this limitation, Bae and Ko (2020) introduced finite mixtures of size-biased distributions with a Weibull baseline, while Bae and Miljkovic (2024) modelled heavy-tailed loss data using finite mixtures of size-biased distributions with truncated log-normal distributions. These studies demonstrate the potential of size-biased distributions and mixture models for capturing the complexity of insurance claim data. A variety of finite mixture models constructed using different mixture components, such as log-normal and Weibull mixtures, have also been investigated by Blostein and Miljkovic (2019).

A critical aspect of insurance claim data is the inherent heterogeneity among policyholders and claims. This heterogeneity can arise from various factors, such as differences in risk characteristics, risk preferences, policy terms, and claim-reporting behaviour (De Jong and Heller 2008; Tseung et al. 2023). While some risk characteristics can be quantifiable through covariates, there are factors that may not be directly observable. Ignoring such unobservable heterogeneity can lead to biased parameter estimates and inaccurate predictions. To address this issue, mixture models have been widely used in insurance modelling to represent the claim distribution as a combination of multiple underlying distributions, each corresponding to a different segment of the insured population (Frees and Valdez 1998). While finite mixture models can effectively account for subpopulation heterogeneity, they often require a large number of mixture components, which can pose challenges for statistical inference and actuarial ratemaking.

To construct a probabilistic structure that enhances predictive performance while supporting fairer risk evaluation across policyholders, we consider a continuous mixture formulation by incorporating a random-effects model with latent variables to capture individual or unit-specific heterogeneity in the probability weighting for large losses.

This paper contributes to the literature by introducing a novel mixed size-biased log-normal (MSBLN) distribution that combines the advantages of size-biased weighting, mixture modelling, and Bayesian ratemaking. The proposed model provides a flexible and robust framework for modelling insurance claim data, accommodating both heavy-tailed

behaviour and individual heterogeneity. We investigate the distributional properties of the MSBLN model, present estimation procedures, and demonstrate its applicability using real insurance loss data.

The remainder of this paper is organized as follows. Section 2 reviews preliminary concepts, including basic distributions, the expectation–maximization (EM) algorithm, and the likelihood ratio test (LRT). Section 3 introduces the mixed size-biased log-normal distribution with a truncated normal prior and examines its key theoretical properties. Estimation via the EM algorithm is presented in Section 4. Section 5 extends the model to a regression setting. A real-data application is provided in Section 6, followed by a discussion of a multivariate extension in Section 7. Section 8 concludes the paper.

2. Preliminaries

2.1. Distributions Used

2.1.1. Log-Normal Distribution

A positive random variable X is log-normally distributed with parameters (μ, σ^2), denoted as $X \sim \text{Lognormal}(\mu, \sigma^2)$ if $\log X \sim N(\mu, \sigma^2)$. Its probability density function (pdf) and cumulative distribution function (cdf) are

$$\begin{aligned} f_X(x) &= \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\}, \qquad x > 0, \ \sigma > 0, \\ F_X(x) &= \Phi\left(\frac{\log x - \mu}{\sigma}\right), \end{aligned}$$

respectively, where Φ denotes the standard normal cumulative distribution function.

2.1.2. Truncated Normal Distribution

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. When observations are restricted to the interval (a, b), the conditional distribution of X on (a, b) is called the truncated normal distribution.

Define the standard normal probability density function and cumulative distribution function by

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \qquad \Phi(z) = \int_{-\infty}^{z} \varphi(t)dt,$$

and the standardized bounds

$$\alpha = \frac{a-\mu}{\sigma}, \qquad \beta = \frac{b-\mu}{\sigma}.$$

Then, the truncated normal probability density function and cumulative distribution function on (a, b), respectively, are

$$f_{TN}(x) = \frac{\varphi\left(\frac{x-\mu}{\sigma}\right)}{\sigma\left[\Phi(\beta) - \Phi(\alpha)\right]}, \qquad a < x < b, \tag{1}$$

$$F_{TN}(x) = \frac{\Phi\left(\frac{x-\mu}{\sigma}\right) - \Phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)}, \qquad a < x < b.$$

The conditional mean is

$$\mathbb{E}[X|a < X < b] = \mu + \sigma \frac{\varphi(\alpha) - \varphi(\beta)}{\Phi(\beta) - \Phi(\alpha)},$$

and the conditional moment generating function is

$$M_X(t) = \mathbb{E}\left[e^{tX} \middle| a < X < b\right] = \exp\left(\mu t + \tfrac{1}{2}\sigma^2 t^2\right) \frac{\Phi(\beta - \sigma t) - \Phi(\alpha - \sigma t)}{\Phi(\beta) - \Phi(\alpha)}. \tag{2}$$

2.2. The Size-Biased Distribution

Let $f(x;\boldsymbol{\theta})$ be a probability density function from a parametric class with parameter(s) $\boldsymbol{\theta}$ for $x \in (0,\infty)$. Then, for a fixed $\lambda > 0$, the pdf of a size-biased distribution with baseline f is given as

$$g(x|\lambda;\boldsymbol{\theta}) = \frac{x^{\lambda} f(x;\boldsymbol{\theta})}{\mu_f(\lambda;\boldsymbol{\theta})},$$

provided that $\mu_f(\lambda;\boldsymbol{\theta}) = \int_0^{\infty} x^{\lambda} f(x;\boldsymbol{\theta})dx < \infty$. Here, the parameter λ is referred to as the order of size-biased weighting. A higher λ value leads to greater probability in the tail; thus, the resulting size-biased distribution can be useful when the data contains large losses. In particular, when $\lambda = 1$, the distribution with pdf g is often referred to as the length-biased version of f.

It is easy to show that the size-biased distribution family on a common baseline is stochastically ordered in terms of the parameter λ in the sense that $\bar{G}(x|\lambda_2) \geq \bar{G}(x|\lambda_1)$ if $\lambda_2 \geq \lambda_1$, where $\bar{G}(x|\lambda) = \int_x^{\infty} g(y|\lambda)dy$ is the survival function of g. This ordering property makes the size-biased weighting a simple but effective method by which to construct new distributions (and their mixtures) when data exhibits significantly heavy (right) tails.

2.3. Estimation Methods: Expectation–Maximization Algorithm

Let Y denote the observed data and Z the latent variables, with complete data $X = (Y, Z)$ and parameter vector θ. The Expectation–Maximization algorithm iteratively maximizes the observed-data log-likelihood $\ell(\theta; Y)$ by alternating the following two steps until convergence.

E-Step:

Compute the Q-function,

$$Q(\theta \mid \theta^{(t)}) = \mathbb{E}_{Z|Y,\theta^{(t)}}\left[\log p(Y, Z \mid \theta)\right],$$

which is the conditional expectation of the complete-data log-likelihood given Y under the current parameter $\theta^{(t)}$.

M-Step:

Update the parameter by maximizing the Q-function:

$$\theta^{(t+1)} = \arg\max_{\theta} \; Q(\theta \mid \theta^{(t)}).$$

Each iteration does not decrease the observed-data log-likelihood $\ell(\theta; Y)$, and the procedure stops when a chosen convergence criterion is met, e.g., when the change in ℓ or in θ falls below tolerance.

2.4. Testing Methods: Likelihood Ratio Test

The likelihood ratio test compares two nested models to evaluate whether the more complex model provides a significantly better fit to the data. Under this test, the null hypothesis (H_0) corresponds to the simpler, restricted model, while the alternative hypothesis (H_1) represents the more general, unrestricted model.

Given a sample Y, let $\hat{\ell}_0 = \log L(\hat{\theta}_0; Y)$ and $\hat{\ell}_1 = \log L(\hat{\theta}_1; Y)$ be the maximum log-likelihoods under two models, respectively. The test statistic, Λ, is then calculated as the difference in their log-likelihoods, scaled by 2:

$$\Lambda = 2(\hat{\ell}_1 - \hat{\ell}_0).$$

This statistic is equivalent to the difference in deviance between the two models. A larger value of Λ indicates a greater improvement in fit provided by the more general model.

3. Mixed Size-Biased Log-Normal with Truncated Normal Prior

The density function of a log-normal distribution with parameters μ and σ is given as

$$f(x;\mu,\sigma) = \frac{1}{\sigma x\sqrt{2\pi}} \exp\left(-\frac{(\log x-\mu)^2}{2\sigma^2}\right), \quad x>0,$$

whose moments $\mu_f(\lambda;\mu,\sigma) = \int_0^\infty x^\lambda f(x;\mu,\sigma)dx = \exp(\mu\lambda+\sigma^2\lambda^2/2)$ exist for all λ.

Then, given $\lambda > 0$, the pdf and the cdf of a size-biased log-normal distribution are expressed as

$$\begin{aligned} g(x|\lambda;\mu,\sigma) &= \frac{x^\lambda f(x;\mu,\sigma)}{\mu_f(\lambda;\mu,\sigma)} \\ &= \exp\left\{-(\mu-\log x)\lambda - \frac{1}{2}\sigma^2\lambda^2\right\}\frac{1}{x\sigma\sqrt{2\pi}}\exp\left\{-\frac{(\log x-\mu)^2}{2\sigma^2}\right\} \\ &= \frac{1}{x\sigma\sqrt{2\pi}}\exp\left\{-\frac{\left(\log x-(\mu+\lambda\sigma^2)\right)^2}{2\sigma^2}\right\}, \end{aligned}$$

and

$$G(x|\lambda;\mu,\sigma) = \Phi\left(\frac{\log x-(\mu+\lambda\sigma^2)}{\sigma}\right),$$

respectively. Here, $\Phi(\cdot)$ denotes the standard normal cdf .

That is, the conditional distribution is log-normal with parameter $\tilde{\mu} = \mu+\lambda\sigma^2$ and $\tilde{\sigma} = \sigma$. The k-th moment is expressed explicitly as

$$\begin{aligned} \mu_f^{(k)}(\lambda;\mu,\sigma) := \mathrm{E}[X^k|\lambda;\mu,\sigma] &= \int_0^\infty x^k g(x|\lambda;\mu,\sigma)\,dx \\ &= \frac{\int_0^\infty x^{k+\lambda}f(x;\mu,\sigma)dx}{\int_0^\infty x^\lambda f(x;\mu,\sigma)dx} \\ &= \frac{\exp\{\mu(\lambda+k)+\sigma^2(\lambda+k)^2/2\}}{\exp\{\mu\lambda+\sigma^2\lambda^2/2\}} \\ &= \exp\left\{(\mu+\sigma^2\lambda)k+\sigma^2k^2/2\right\}. \end{aligned}$$

We now consider a hierarchical model by treating λ as an unobservable latent variable. The randomization of λ, i.e., the order of size-biased weighting, aims to capture individual or unit-specific heterogeneity in the probability weighting for large losses.

Specifically, we assume λ has a truncated normal distribution on $(0,\infty)$, or a normal distribution left truncated at 0, with pdf

$$\pi(\lambda;\mu_0,\sigma_0) = \frac{\phi\left(\frac{\lambda-\mu_0}{\sigma_0}\right)}{\sigma_0\Phi\left(\frac{\mu_0}{\sigma_0}\right)} = \frac{\exp\left\{-\frac{1}{2\sigma_0^2}(\lambda-\mu_0)^2\right\}}{\sigma_0\sqrt{2\pi}\Phi\left(\frac{\mu_0}{\sigma_0}\right)}, \quad \lambda>0. \tag{3}$$

Note that (3) is a special case of (1), with $a=0$ and $b=\infty$, so $\alpha = -\mu_0/\sigma_0$ and $\beta=\infty$; hence, the normalizing constant becomes $\Phi(\beta)-\Phi(\alpha) = 1-\Phi(-\mu_0/\sigma_0) = \Phi(\mu_0/\sigma_0)$.

By (2), the moment generating function of the latent variable Λ is

$$M_\Lambda(t) = \mathrm{E}[e^{t\Lambda};\mu_0,\sigma_0] = \exp\left\{\mu_0 t+\frac{1}{2}\sigma_0^2t^2\right\}\frac{\Phi(\mu_0/\sigma_0+\sigma_0 t)}{\Phi(\mu_0/\sigma_0)}.$$

Using this, it is easy to find the following expressions for the first and second moment of Λ:

$$\mathrm{E}[\Lambda;\mu_0,\sigma_0] = \mu_0 + \sigma_0 \frac{\phi(\mu_0/\sigma_0)}{\Phi(\mu_0/\sigma_0)} \tag{4}$$

$$\mathrm{E}[\Lambda^2;\mu_0,\sigma_0] = \mu_0^2 + \sigma_0^2 + \mu_0\sigma_0 \frac{\phi(\mu_0/\sigma_0)}{\Phi(\mu_0/\sigma_0)}. \tag{5}$$

By completing squares and using the cdf of standard normal distribution, it is easy to obtain the following expression for the pdf of the unconditional (or marginal) distribution:

$$\begin{aligned} & g(x;\mu,\sigma,\mu_0,\sigma_0) \\ &= \int_0^\infty g(x|\lambda;\mu,\sigma)\pi(\lambda;\mu_0,\sigma_0)\,d\lambda \\ &= \int_0^\infty \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\log x-\mu-\lambda\sigma^2)^2}{2\sigma^2}\right\} \frac{\exp\left\{-\frac{1}{2\sigma_0^2}(\lambda-\mu_0)^2\right\}}{\sigma_0\sqrt{2\pi}\Phi\left(\frac{\mu_0}{\sigma_0}\right)}\,d\lambda \\ &= \frac{1}{x\sigma\sqrt{(1+\sigma^2\sigma_0^2)2\pi}} \exp\left\{-\frac{(\log x-\mu-\mu_0\sigma^2)^2}{2\sigma^2(1+\sigma^2\sigma_0^2)}\right\} \frac{\Phi\left(\frac{(\log x-\mu)\sigma_0+\frac{\mu_0}{\sigma_0}}{\sqrt{1+\sigma^2\sigma_0^2}}\right)}{\Phi\left(\frac{\mu_0}{\sigma_0}\right)}. \end{aligned} \tag{6}$$

The cdf of the marginal distribution is

$$\begin{aligned} G(x;\mu,\sigma,\mu_0,\sigma_0) &= \int_0^\infty G(x|\lambda;\mu,\sigma)\pi(\lambda;\mu_0,\sigma_0)\,d\lambda \\ &= \frac{1}{\sigma_0\Phi\left(\frac{\mu_0}{\sigma_0}\right)} \int_0^\infty \Phi\left(-\sigma\lambda+\frac{\log x-\mu}{\sigma}\right)\phi\left(\frac{\lambda-\mu_0}{\sigma_0}\right)d\lambda. \end{aligned} \tag{7}$$

Since $\Phi\left(-\sigma\lambda+\frac{\log x-\mu}{\sigma}\right) \le \Phi\left(\frac{\log x-\mu}{\sigma}\right)$, for any $\lambda > 0$, we can see

$$\begin{aligned} G(x;\mu,\sigma,\mu_0,\sigma_0) &\le \frac{\Phi\left(\frac{\log x-\mu}{\sigma}\right)}{\sigma_0\Phi\left(\frac{\mu_0}{\sigma_0}\right)} \int_0^\infty \phi\left(\frac{\lambda-\mu_0}{\sigma_0}\right)d\lambda \\ &= \Phi\left(\frac{\log x-\mu}{\sigma}\right) = F(x;\mu,\sigma), \end{aligned}$$

which implies that the tail of the MSBLN distribution with a truncated normal prior (MSBLN-TN) is heavier than that of the baseline log-normal distribution. The following lemma is useful in providing a convenient expression for the cdf (7).

Lemma 1. *Let $\Phi_2(x,y;\rho) = P(X \le x, Y \le y)$ be the joint cdf of standard bivariate normal random variables with the correlation coefficient ρ. For constants a, b, and c,*

$$\int_c^\infty \Phi(a+bz)\phi(z)dz = \Phi_2\left(-c, \frac{a}{\sqrt{1+b^2}}; \rho = \frac{b}{\sqrt{1+b^2}}\right). \tag{8}$$

Proof. See Appendix A. □

By Lemma 1, cdf (7) can be expressed as follows:

$$\begin{aligned} G(x;\mu,\sigma,\mu_0,\sigma_0) &= \frac{1}{\sigma_0\Phi\left(\frac{\mu_0}{\sigma_0}\right)} \int_0^\infty \Phi\left(-\sigma\lambda+\frac{\log x-\mu}{\sigma}\right)\phi\left(\frac{\lambda-\mu_0}{\sigma_0}\right)d\lambda \\ &= \frac{1}{\Phi\left(\frac{\mu_0}{\sigma_0}\right)} \int_c^\infty \Phi(a+bz)\phi(z)\,d\lambda \\ &= \frac{\Phi_2\left(-c, \frac{a}{\sqrt{1+b^2}}; \rho = \frac{b}{\sqrt{1+b^2}}\right)}{\Phi\left(\frac{\mu_0}{\sigma_0}\right)}, \end{aligned} \tag{9}$$

where $a = (\log x - \mu - \sigma^2\mu_0)/\sigma$, $b = -\sigma\sigma_0$, and $c = -\mu_0/\sigma_0$.

From (9), we can see that the MSBLN-TN distribution is characterized by the following expressions involving model parameters: $(\mu_0/\sigma_0, \sigma\sigma_0, \mu + \sigma^2\mu_0, \sigma\sqrt{1+\sigma^2\sigma_0^2})$. It is easy to see that the mapping from $(\mu, \sigma, \mu_0, \sigma_0)$ to $(\mu_0/\sigma_0, \sigma\sigma_0, \mu + \sigma^2\mu_0, \sigma\sqrt{1+\sigma^2\sigma_0^2})$ is bijective; thus, the model is identifiable. More formally, the MSBLN-TN model is identifiable if

$$G(x; \mu, \sigma, \mu_0, \sigma_0) = G(x; \mu', \sigma', \mu_0', \sigma_0') \tag{10}$$

for all x, implying $(\mu, \sigma, \mu_0, \sigma_0) = (\mu', \sigma', \mu_0', \sigma_0')$. Through the identifiability of the standard bivariate normal distribution in (9), the condition (10) implies

$$\begin{cases} \mu_0/\sigma_0 = \mu_0'/\sigma_0' \\ \sigma\sigma_0 = \sigma'\sigma_0' \\ \left(\frac{1}{\sigma} - \frac{1}{\sigma'}\right)\log x = \left(\frac{\mu}{\sigma} - \frac{\mu'}{\sigma'}\right) + \sigma\mu_0 - \sigma'\mu_0', \end{cases}$$

for all x. Clearly, we have $\sigma = \sigma', \sigma_0 = \sigma_0', \mu = \mu', \mu_0 = \mu_0'$; thus, the model parameters in the MSBLN-TN distribution are identifiable.

By using the moment generating function of the truncated normal distribution, it is easy to obtain the following expression for the k-th moment of the unconditional distribution.

$$\begin{aligned} \mathrm{E}[X^k; \mu, \sigma, \mu_0, \sigma_0] &= \int_0^\infty \mathrm{E}[X^k|\lambda; \mu, \sigma]\pi(\lambda; \mu_0, \sigma_0)\, d\lambda \\ &= \exp\{\mu k + \sigma^2 k^2/2\} \int_0^\infty e^{\sigma^2 k\lambda}\pi(\lambda; \mu_0, \sigma_0)\, d\lambda \\ &= \exp\{(\mu + \sigma^2\mu_0)k + (\sigma^2 + \sigma^4\sigma_0^2)k^2/2\} \frac{\Phi\left(\frac{\mu_0}{\sigma_0} + \sigma^2\sigma_0 k\right)}{\Phi\left(\frac{\mu_0}{\sigma_0}\right)}. \end{aligned} \tag{11}$$

3.1. Tail Risk Measures

In actuarial science and related applications, tail risk measures play a key role in assessing risk exposure to extreme events. Here, we consider the two most popular tail risk measures in practice: Value at Risk (VaR) and Tail Value at Risk (TVaR). The VaR of risk X at level α is defined as the α-quantile of X. Let X have the MSBLN-TN distribution with the cdf $G(x; \mu, \sigma, \mu_0, \sigma_0)$.

$$\mathrm{VaR}_\alpha(X) = \pi_\alpha = G^{-1}(\alpha; \mu, \sigma, \mu_0, \sigma_0),$$

or equivalent, π_α, solves the equation

$$\frac{\Phi_2\left(\mu_0/\sigma_0, \frac{(\log\pi_\alpha - \mu)/\sigma - \sigma\mu_0}{\sqrt{1+\sigma^2\sigma_0^2}}; \rho = \frac{-\sigma\sigma_0}{\sqrt{1+\sigma^2\sigma_0^2}}\right)}{\Phi\left(\frac{\mu_0}{\sigma_0}\right)} = \alpha. \tag{12}$$

The TVaR of a risk X at level α, denoted by $\mathrm{TVaR}_\alpha(X)$, is defined as the conditional expectation of X given that X is larger than the VaR. Specifically, we obtain the following convenient expression for the TVaR of MSBLN-TN random variable X.

Proposition 1.

$$\begin{aligned} TVaR_\alpha(X) &= E[X \,|\, X > \pi_\alpha] \\ &= \frac{\exp\left[\mu + \sigma^2/2 + \mu_0\sigma^2 + \sigma_0^2\sigma^4/2\right]\Phi_2\left(-c, \frac{a}{\sqrt{1+b^2}}; \rho = \frac{b}{\sqrt{1+b^2}}\right)}{(1-\alpha)\Phi\left(\frac{\mu_0}{\sigma_0}\right)}, \end{aligned} \tag{13}$$

where $a = \sigma(\mu_0 + \sigma_0^2\sigma^2 + 1) + (\mu - \log \pi_\alpha)/\sigma$, $b = \sigma_0\sigma$, *and* $c = -(\mu_0 + \sigma_0^2\sigma^2)/\sigma_0$.

Proof. See Appendix B. □

3.2. Bayesian Analysis

Suppose a random sample $\boldsymbol{x} = (x_1, \ldots, x_n)$ of size n is observed. Here, we assume that the observations are independent given a realization of the mixing random variable Λ. Then, the posterior density of Λ with a truncated normal prior distribution is

$$\begin{aligned}
\pi(\lambda|\boldsymbol{x};\mu,\sigma,\mu_0,\sigma_0) &\propto g(\boldsymbol{x}|\lambda;\mu,\sigma)\pi(\lambda;\mu_0,\sigma_0)\\
&\propto \left(\prod_{i=1}^{n} f(x_i|\lambda;\mu,\sigma)\right)\pi(\lambda;\mu_0,\sigma_0)\\
&\propto \exp\left\{\left(\sum_{i=1}^{n}\log x_i - n\mu\right)\lambda - \frac{n}{2}\sigma^2\lambda^2\right\}\exp\left\{-\frac{1}{2\sigma_0^2}(\lambda-\mu_0)^2\right\}\\
&\propto \exp\left\{-\frac{1}{2}(n\sigma^2+\sigma_0^{-2})\left(\lambda - \frac{\sum_{i=1}^{n}\log x_i - n\mu + \mu_0/\sigma_0^2}{n\sigma^2+\sigma_0^{-2}}\right)^2\right\}. \qquad (14)
\end{aligned}$$

Thus, the posterior distribution of Λ is a truncated normal distribution with parameters

$$\begin{aligned}
\mu_0^* &= \frac{\sum_{i=1}^{n}\log x_i - n\mu + \mu_0/\sigma_0^2}{n\sigma^2+\sigma_0^{-2}} = \frac{\sigma_0^2(\sum_{i=1}^{n}\log x_i - n\mu) + \mu_0}{1+n\sigma^2\sigma_0^2}\\
\sigma_0^* &= \frac{1}{\sqrt{n\sigma^2+\sigma_0^{-2}}} = \frac{\sigma_0}{\sqrt{1+n\sigma^2\sigma_0^2}}.
\end{aligned}$$

In other words, a truncated normal distribution is a conjugate prior of the size-biased log-normal distribution.

Similarly as in (6), the predictive density of X_{n+1}, given $\boldsymbol{x}$, is expressed as

$$\begin{aligned}
&g(x\,|\,\boldsymbol{x},\mu,\sigma,\mu_0,\sigma_0)\\
&= \int_0^\infty g(x|\lambda;\mu,\sigma)\pi(\lambda|\boldsymbol{x};\mu,\sigma,\mu_0,\sigma_0)\,d\lambda\\
&= \int_0^\infty \frac{1}{x\sigma\sqrt{2\pi}}\exp\left\{-\frac{(\log x-\mu-\lambda\sigma^2)^2}{2\sigma^2}\right\}\frac{\exp\left\{-\frac{1}{2\sigma_0^{*2}}(\lambda-\mu_0^*)^2\right\}}{\sigma_0^*\sqrt{2\pi}\Phi\left(\frac{\mu_0^*}{\sigma_0^*}\right)}\,d\lambda\\
&= \frac{1}{x\sigma\sqrt{(1+\sigma^2\sigma_0^{*2})2\pi}}\exp\left\{-\frac{(\log x-\mu-\mu_0^*\sigma^2)^2}{2\sigma^2(1+\sigma^2\sigma_0^{*2})}\right\}\frac{\Phi\left(\frac{(\log x-\mu)\sigma_0^*+\frac{\mu_0^*}{\sigma_0^*}}{\sqrt{1+\sigma^2\sigma_0^{*2}}}\right)}{\Phi\left(\frac{\mu_0^*}{\sigma_0^*}\right)}. \qquad (15)
\end{aligned}$$

In an actuarial context, the expected value of X, given the observations $\boldsymbol{x}$, referred to as the Bayesian premium, is important. More generally, the k-th moment of the predictive distribution is expressed as

$$\begin{aligned}
&\mathrm{E}[X^k|\boldsymbol{x};\mu,\sigma,\mu_0,\sigma_0]\\
&= \exp\{(\mu+\sigma^2\mu_0^*)k + (\sigma^2+\sigma^4\sigma_0^{*2})k^2/2\}\frac{\Phi\left(\frac{\mu_0^*}{\sigma_0^*}+\sigma^2\sigma_0^*k\right)}{\Phi\left(\frac{\mu_0^*}{\sigma_0^*}\right)}\\
&= \exp\left\{\left[\mu+\left(\frac{\sigma^2\sigma_0^2}{1+n\sigma^2\sigma_0^2}\right)\left(\sum_{i=1}^{n}\log x_i - n\mu + \frac{\mu_0}{\sigma_0^2}\right)\right]k + \left[\frac{(n+1)\sigma^4\sigma_0^2+\sigma^2}{2(1+n\sigma^2\sigma_0^2)}\right]k^2\right\}\\
&\times\frac{\Phi\left(\frac{\sigma_0(\sum_{i=1}^{n}\log x_i - n\mu + k\sigma^2)+\mu_0/\sigma_0}{\sqrt{n\sigma^2+\sigma_0^{-2}}}\right)}{\Phi\left(\frac{\sigma_0(\sum_{i=1}^{n}\log x_i - n\mu)+\mu_0/\sigma_0}{\sqrt{1+n\sigma^2\sigma_0^2}}\right)}.
\end{aligned}$$

4. Parameter Estimation

Let $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ be a random sample of size n drawn from the proposed model with the pdf (6), which involves four unknown parameters: μ, σ, μ_0, and σ_0. We further denote the set of model parameters as $\boldsymbol{\theta} = (\mu, \sigma, \mu_0, \sigma_0)$. For each observation x_i, the unobservable latent random variable Λ_i may be treated as a missing value; thus, we proceed with the EM algorithm (Dempster et al. 1977) to obtain the maximum likelihood estimates of $\boldsymbol{\theta}$ numerically. As briefly described in Section 2.3, the EM algorithm updates the parameter estimates by maximizing the expectation of the complete-data log-likelihood function, often referred to as the Q function, with respect to the conditional distribution of the latent variables given the data and the current estimates, or in our case, the posterior distribution of $\Lambda_i, i = 1, \ldots, n$.

Formally, the complete-data log-likelihood function is given as

$$\begin{aligned} \log L(\boldsymbol{\theta} \,|\, \boldsymbol{x}, \Lambda) &= \log\left(\prod_{i=1}^{n} g(x_i|\Lambda_i; \mu, \sigma)\pi(\Lambda_i; \mu_0, \sigma_0)\right) \\ &= \sum_{i=1}^{n} \log g(x_i|\Lambda_i; \mu, \sigma) + \sum_{i=1}^{n} \log \pi(\Lambda_i; \mu_0, \sigma_0). \end{aligned}$$

For the E-step of the EM algorithm, the expectation of the complete-data log-likelihood function with respect to the posterior distribution of Λ, given the observations $\boldsymbol{x}$ and the current parameter estimates $\boldsymbol{\theta}^{(s)}$, denoted by $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s)})$ and referred to as the Q-function, is expressed (up to a constant) as

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s)}) &= \sum_{i=1}^{n} \mathrm{E}_{\Lambda_i|x_i,\boldsymbol{\theta}^{(s)}}[\log g(x_i|\Lambda_i; \mu, \sigma)] + \sum_{i=1}^{n} \mathrm{E}_{\Lambda_i|x_i,\boldsymbol{\theta}^{(s)}}[\log \pi(\Lambda_i; \mu_0, \sigma_0)] \\ &= (A) + (B), \end{aligned}$$

where

$$\begin{aligned} (A) &= -n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\log x_i - \mu)^2 - \mu\sum_{i=1}^{n}\mathrm{E}[\Lambda_i|x_i, \boldsymbol{\theta}^{(s)}] - \frac{\sigma^2}{2}\sum_{i=1}^{n}\mathrm{E}[\Lambda_i^2|x_i, \boldsymbol{\theta}^{(s)}] \\ (B) &= -\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}\mathrm{E}[(\Lambda_i - \mu_0)^2|x_i, \boldsymbol{\theta}^{(s)}] - n\log\sigma_0 - n\log\Phi(\mu_0/\sigma_0). \end{aligned}$$

With the truncated normal posterior distribution being as given in Section 3.2, we have

$$\begin{aligned} \mathrm{E}[\Lambda_i|x_i, \boldsymbol{\theta}^{(s)}] &= \mu_{0i}^{*(s)} + \sigma_0^{*(s)}\frac{\phi(\mu_{0i}^{*(s)}/\sigma_0^{*(s)})}{\Phi(\mu_{0i}^{*(s)}/\sigma_0^{*(s)})} \\ \mathrm{E}[\Lambda_i^2|x_i, \boldsymbol{\theta}^{(s)}] &= (\sigma_0^{*(s)})^2 + (\mu_{0i}^{*(s)})^2 + \sigma_0^{*(s)}\mu_{0i}^{*(s)}\frac{\phi(\mu_{0i}^{*(s)}/\sigma_0^{*(s)})}{\Phi(\mu_{0i}^{*(s)}/\sigma_0^{*(s)})}, \end{aligned}$$

where

$$\begin{aligned} \mu_{0i}^{*(s)} &= \frac{\mu_0^{(s)} - \mu^{(s)}(\sigma_0^{(s)})^2 + (\sigma_0^{(s)})^2\log x_i}{1 + (\sigma^{(s)})^2(\sigma_0^{(s)})^2}, i = 1, \ldots, n, \\ \sigma_0^{*(s)} &= \frac{\sigma_0^{(s)}}{\sqrt{1 + (\sigma^{(s)})^2(\sigma_0^{(s)})^2}}. \end{aligned}$$

Due to the separation of the Q-function into (A) and (B), the M-step of the EM algorithm can proceed separately for updating the estimates of (μ, σ) and (μ_0, σ_0). In fact, the optimization problem (A) provides the following explicit solution:

$$\begin{aligned}(\sigma^{(s+1)})^2 &= \frac{1}{2(\bar{b}-\bar{a}^2)}\left(-1+\sqrt{1+4(\bar{b}-\bar{a}^2)\left(n^{-1}\sum_{i=1}^{n}(\log x_i)^2-\bar{l}^2\right)}\right)\\ \mu^{(s+1)} &= \bar{l}-\bar{a}(\sigma^{(s+1)})^2,\end{aligned}$$

where $\bar{a}=n^{-1}\sum_{i=1}^{n}\mathrm{E}[\Lambda_i|x_i,\boldsymbol{\theta}^{(s)}]$, $\bar{b}=n^{-1}\sum_{i=1}^{n}\mathrm{E}[\Lambda_i^2|x_i,\boldsymbol{\theta}^{(s)}]$, and $\bar{l}=n^{-1}\sum_{i=1}^{n}\log x_i$. The updated parameters $\mu_0^{(s+1)}$ and $\sigma_0^{(s+1)}$ are the solutions of the following system of equations:

$$\begin{cases}\mu_0+\sigma_0\frac{\phi(\mu_0/\sigma_0)}{\Phi(\mu_0/\sigma_0)}=\bar{a}\\ \bar{a}\mu_0+\sigma_0^2=\bar{b},\end{cases}$$

with the constraint $\mu_0<\bar{b}/\bar{a}$. For the initialization of the EM algorithm, we use a grid-based approach in setting various sets of starting values, and the one that yields the largest likelihood is selected for the final fitted model. The estimation procedure is summarized in Appendix D, Algorithm A1.

5. Regression Model

In this section, we consider a regression model where the expected value of a mixed size-biased log-normal response variable Y_i, denoted by μ_{Y_i}, is a function of the predictor variables $X_{i1},\ldots,X_{ip}$:

$$\psi(\mu_{Y_i})=\beta_0+\beta_1X_{i1}+\ldots+\beta_pX_{ip},$$

for a link function ψ. From (11), the (unconditional) expectation of Y_i is given as

$$\mu_{Y_i}:=\mathrm{E}[Y_i;\mu_i,\sigma,\mu_0,\sigma_0]=\exp\{\mu_i+(\mu_0+1/2)\sigma^2+\sigma^4\sigma_0^2/2\}\frac{\Phi\left(\frac{\mu_0}{\sigma_0}+\sigma^2\sigma_0\right)}{\Phi\left(\frac{\mu_0}{\sigma_0}\right)}.$$

Therefore, with the log-link function $\psi(x)=\log(x)$, we have

$$\begin{aligned}\psi(\mu_{Y_i}) &= \mu_i+(\mu_0+1/2)\sigma^2+\sigma^4\sigma_0^2/2+\log\left[\Phi\left(\frac{\mu_0}{\sigma_0}+\sigma^2\sigma_0\right)/\Phi\left(\frac{\mu_0}{\sigma_0}\right)\right]\\ &= \beta_0+\beta_1X_{i1}+\ldots+\beta_pX_{ip}.\end{aligned}$$

From this, it is natural to specify the regression model as

$$\begin{aligned}\mu_i &= \beta_1X_{i1}+\ldots+\beta_pX_{ip}=X_i^T\boldsymbol{\beta}\\ \beta_0 &= (\mu_0+1/2)\sigma^2+\sigma^4\sigma_0^2/2+\log\left[\Phi\left(\frac{\mu_0}{\sigma_0}+\sigma^2\sigma_0\right)/\Phi\left(\frac{\mu_0}{\sigma_0}\right)\right].\end{aligned}$$

Since β_0 is given as a function of other model parameters, i.e., σ, μ_0, and σ_0, it is not identifiable; thus, it is excluded from estimation.

The maximum likelihood estimation method can be used to estimate regression coefficients $\boldsymbol{\beta}^T=(\beta_1,\ldots,\beta_p)$ along with the other model parameters $\boldsymbol{\theta}=(\sigma,\mu_0,\sigma_0)$. Similarly as in Section 4, we apply the EM algorithm for the estimation of the proposed regression model.

The complete-data log-likelihood function is given as

$$\begin{aligned}\log L(\boldsymbol{\beta},\boldsymbol{\theta}\,|\,\boldsymbol{y},\boldsymbol{x},\Lambda) &= \log\left(\prod_{i=1}^{n}g(y_i|\Lambda_i;\boldsymbol{x}_i^T\boldsymbol{\beta},\sigma)\pi(\Lambda_i;\mu_0,\sigma_0)\right)\\ &= \sum_{i=1}^{n}\log g(y_i|\Lambda_i;\boldsymbol{x}_i^T\boldsymbol{\beta},\sigma)+\sum_{i=1}^{n}\log\pi(\Lambda_i;\mu_0,\sigma_0).\end{aligned}$$

For the E-step of the EM algorithm, the Q-function is expressed as

$$\begin{aligned} & Q((\boldsymbol{\beta},\boldsymbol{\theta})|(\boldsymbol{\beta}^{(s)},\boldsymbol{\theta}^{(s)})) \\ & \quad = \sum_{i=1}^{n} \mathrm{E}_{\Lambda_i|y_i,\boldsymbol{x}_i^T\boldsymbol{\beta}^{(s)},\boldsymbol{\theta}^{(s)}}[\log g(y_i|\Lambda_i;\mu_i,\sigma)] + \sum_{i=1}^{n} \mathrm{E}_{\Lambda_i|y_i,\boldsymbol{x}_i^T\boldsymbol{\beta}^{(s)},\boldsymbol{\theta}^{(s)}}[\log \pi(\Lambda_i;\mu_0,\sigma_0)] \\ & \quad = (A)+(B), \end{aligned}$$

where

$$\begin{aligned} (A) &= -n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\Big(\log y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\Big)^2 - \sum_{i=1}^{n}\mathrm{E}[\Lambda_i|y_i,\boldsymbol{x}_i^T\boldsymbol{\beta}^{(s)},\boldsymbol{\theta}^{(s)}]\boldsymbol{x}_i^T\boldsymbol{\beta} \\ &\quad -\frac{\sigma^2}{2}\sum_{i=1}^{n}\mathrm{E}[\Lambda_i^2|y_i,\boldsymbol{x}_i^T\boldsymbol{\beta}^{(s)},\boldsymbol{\theta}^{(s)}], \\ (B) &= -\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}\mathrm{E}[(\Lambda_i-\mu_0)^2|y_i,\boldsymbol{x}_i^T\boldsymbol{\beta}^{(s)},\boldsymbol{\theta}^{(s)}] - n\log\sigma_0 - n\log\Phi(\mu_0/\sigma_0). \end{aligned}$$

For each $i = 1,\ldots,$ the posterior moments of the latent variable, given the current parameter estimates, are

$$\begin{aligned} \mathrm{E}[\Lambda_i|y_i,\boldsymbol{x}_i^T\boldsymbol{\beta}^{(s)},\boldsymbol{\theta}^{(s)}] &= \mu_{0i}^{*(s)} + \sigma_0^{*(s)}\frac{\phi(\mu_{0i}^{*(s)}/\sigma_0^{*(s)})}{\Phi(\mu_{0i}^{*(s)}/\sigma_0^{*(s)})}, \\ \mathrm{E}[\Lambda_i^2|y_i,\boldsymbol{x}_i^T\boldsymbol{\beta}^{(s)},\boldsymbol{\theta}^{(s)}] &= (\sigma_0^{*(s)})^2 + (\mu_{0i}^{*(s)})^2 + \sigma_0^{*(s)}\mu_{0i}^{*(s)}\frac{\phi(\mu_{0i}^{*(s)}/\sigma_0^{*(s)})}{\Phi(\mu_{0i}^{*(s)}/\sigma_0^{*(s)})}, \end{aligned}$$

where

$$\begin{aligned} \mu_{0i}^{*(s)} &= \frac{\mu_0^{(s)} - \boldsymbol{x}_i^T\boldsymbol{\beta}^{(s)}(\sigma_0^{(s)})^2 + (\sigma_0^{(s)})^2\log y_i}{1+(\sigma^{(s)})^2(\sigma_0^{(s)})^2} \\ \sigma_0^{*(s)} &= \frac{\sigma_0^{(s)}}{\sqrt{1+(\sigma^{(s)})^2(\sigma_0^{(s)})^2}}. \end{aligned}$$

In the M-step of the EM algorithm, we can find the following explicit solution for the maximization of part (A) of the Q-function.

$$\begin{aligned} (\sigma^{(s+1)})^2 &= \frac{1}{2(\bar{b}-\bar{f})}\left(-1+\sqrt{1+4(\bar{b}-\bar{f})n^{-1}\sum_{i=1}^{n}(\log y_i - d_i)^2}\right) \\ \boldsymbol{\beta}^{(s+1)} &= \left(\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^T\right)^{-1}\Big(\boldsymbol{\zeta}-\sigma^2\boldsymbol{\eta}\Big), \end{aligned}$$

where

$$\begin{aligned} \bar{b} &= n^{-1}\sum_{i=1}^{n}\mathrm{E}[\Lambda_i^2|y_i,\boldsymbol{x}_i^T\boldsymbol{\beta}^{(s)},\boldsymbol{\theta}^{(s)}], \quad \bar{f} = n^{-1}\sum_{i=1}^{n}e_i^2, \\ e_i &= \boldsymbol{x}_i^T\left(\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^T\right)^{-1}\boldsymbol{\eta}, \quad d_i = \boldsymbol{x}_i^T\left(\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^T\right)^{-1}\boldsymbol{\zeta}, \\ \boldsymbol{\zeta} &= \sum_{i=1}^{n}(\log y_i)\boldsymbol{x}_i, \quad \boldsymbol{\eta} = \sum_{i=1}^{n}\mathrm{E}[\Lambda_i|y_i,\boldsymbol{x}_i^T\boldsymbol{\beta}^{(s)},\boldsymbol{\theta}^{(s)}]\boldsymbol{x}_i. \end{aligned}$$

The maximization of part (B) of the Q-function renders the updated estimate for (μ_0,σ_0), which is the solution of the following system of equations:

$$\begin{cases} \mu_0 + \sigma_0 \frac{\phi(\mu_0/\sigma_0)}{\Phi(\mu_0/\sigma_0)} = \bar{a} \\ \bar{a}\mu_0 + \sigma_0^2 = \bar{b}, \end{cases}$$

with the constraint $\mu_0 < \bar{b}/\bar{a}$, where $\bar{a} = n^{-1} \sum_{i=1}^{n} \mathrm{E}[\Lambda_i | y_i, \boldsymbol{x}_i^T \boldsymbol{\beta}^{(s)}, \boldsymbol{\theta}^{(s)}]$. For initialization of the EM-algorithm, we set the regression parameter $\boldsymbol{\beta}^{(0)} = \mathbf{0}$ and use a grid-based approach in setting various sets of plausible starting values. The final fitted model is the one that yields the largest likelihood. The estimation procedure is summarized in Appendix D, Algorithm A2.

6. Application

6.1. Loss Severity Modelling

We demonstrate the application of the MSBLN-TN distribution, estimated via the EM algorithm (Algorithm A1 in Appendix D), to a real loss amount per accident data. A dataset of $n = 3000$ observations has been randomly drawn from a Korean Automobile Insurance company's daily claim data for the period 2007–2017.

The loss amounts were converted to units of 10,000 KRW for analysis. The basic summary statistics are given in Table 1.

Table 1. Summary of data.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Skewness	ex.Kurtosis
58.47	690.75	1110.85	1840.85	2103.01	36,486.64	5.02	46.44

The summary statistics show that the data exhibits strong right-skewness and heavy (right)-tailedness. Even after applying a logarithmic transformation, the distribution still shows right-skewness, as shown in the histogram (Figure 1).

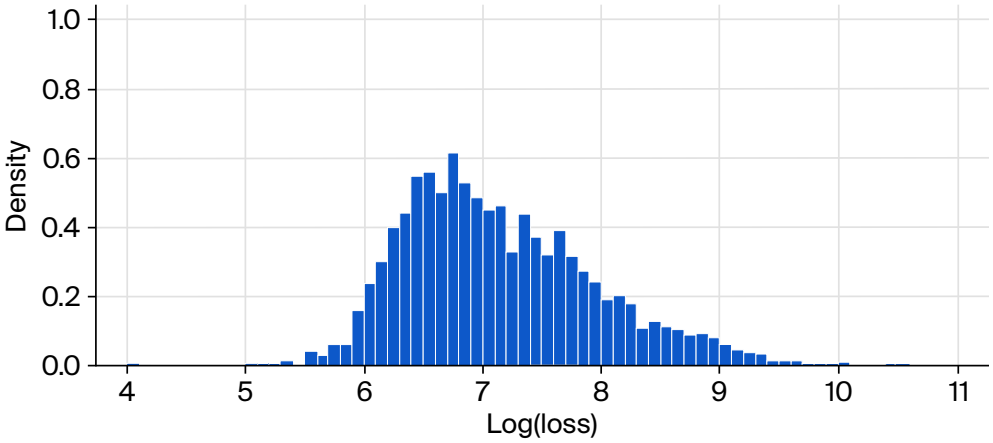


Figure 1. Histogram of log-transformed loss data.

To evaluate the relative goodness of fit of the MSBLN-TN distribution with other heavy-tailed distributions, we consider the following models frequently employed in the loss modelling literature: Pareto, Weibull, log-normal, transformed beta (Trbeta), and Beta prime distributions. We follow Klugman et al. (2019) for the definition and parameterization of Pareto, Weibull, log-normal, and transformed beta distributions. The density of the Beta prime distribution is given as follows:

$$f(x; \alpha, \beta, \sigma) = \frac{(x/\sigma)^{\alpha-1}(1 + (x/\sigma))^{-\alpha-\beta}}{\sigma B(\alpha, \beta)},$$

where $B(\cdot, \cdot)$ is the beta function.

The model selection is based on the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), which are defined as

$$\text{AIC} = -2\log L + 2k, \quad \text{BIC} = -2\log L + k\log n,$$

with L being the likelihood $L(\boldsymbol{\theta} \mid \boldsymbol{x}, \Lambda)$, k being the number of estimated parameters, and n being the sample size. The AIC reflects the goodness of fit through the maximized log-likelihood, while penalising model complexity through the term $2k$. The BIC employs a penalty term of $k \log n$ that depends on sample size, leading to a stronger penalty than the AIC as the sample size increases.

Table 2 presents the parameter estimates, log-likelihood (loglike), AIC, and BIC values for the candidate heavy-tailed distributions. For the estimation of the MSBLN-TN model, we use various sets of starting values from combinations of $\mu \in \{-1, 0, 1\}, \sigma \in \{0.5, 1\}$, $\mu_0 \in \{-1, 0, 1\}$, and $\sigma_0 \in \{0.5, 1\}$. For each set of starting values, the iterative procedure stops when the improvement in the log-likelihood (or the stopping criterion) is less than $\epsilon = 10^{-4}$. The values in parentheses are the bootstrap standard errors of the estimates based on 500 replications. Among all the models, the MSBLN-TN distribution achieves the highest log-likelihood value of $-24{,}892.69$ and the lowest AIC and BIC values—49,793.38 and 49,817.41, respectively. The AIC and BIC differences between MSBLN-TN and the next best model, Beta prime, are 27.08 and 21.07, respectively, while the gaps with respect to the other distributions are substantially larger. Therefore, the MSBLN-TN model can be regarded as the most appropriate model in terms of relative goodness of fit.

Table 2. Comparison of fitted models.

Model	Estimate	Loglike	AIC	BIC
MSBLN-TN $(\mu, \sigma, \mu_0, \sigma_0)$	6.317 (0.014), 0.440 (0.003), -3.835 (1.255), 6.780 (0.397)	$-24{,}892.7$	49,793.4	49,817.4
Pareto (α, θ)	0.00005, 11.674	$-25{,}531.1$	51,066.2	51,078.2
Weibull (τ, θ)	1.109, 1928.630	$-25{,}522.0$	51,047.9	51,059.9
Log-normal (μ, σ)	7.142, 0.803	$-25{,}022.9$	50,049.8	50,061.8
Trbeta $(\alpha, \gamma, \tau, \theta)$	2.932, 0.775, 108.683, 9.635	$-24{,}918.8$	49,845.7	49,869.7
Beta prime (α, β, σ)	3.844, 0.720, 3.765	$-24{,}907.2$	49,820.5	49,838.5

Figure 2 illustrates the fitted probability density function of the MSBLN-TN distribution (red curve) superimposed on the histogram of the log-transformed loss data. The fitted curve closely follows the peak of the empirical distribution and adequately captures the gradual decline in the right tail, indicating a satisfactory fit to the data. Figure 3 shows the P-P plot comparing the empirical cumulative probabilities and those obtained from the fitted MSBLN-TN distribution. The points lie closely along the 45-degree reference line (red line), suggesting a strong agreement between the empirical and fitted distributions.

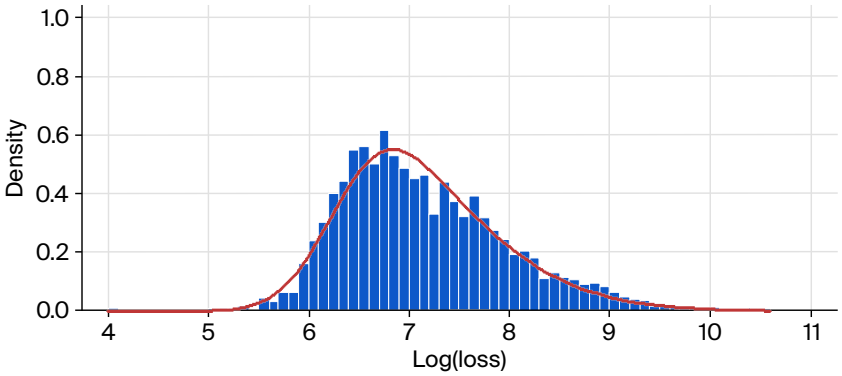


Figure 2. Histogram of log-transformed data with the fitted MSBLN-TN density curve.

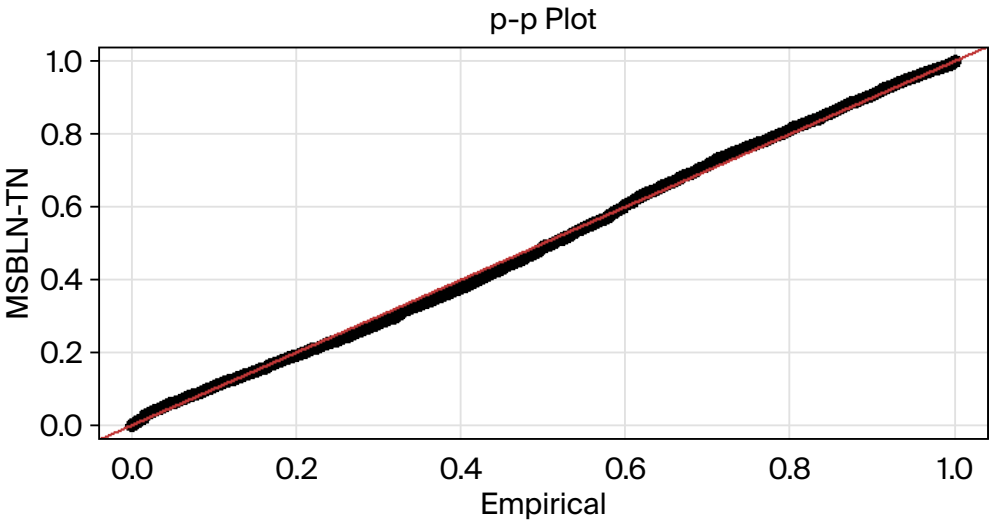


Figure 3. P-P plot of the fitted MSBLN-TN distribution to the empirical distribution.

Table 3 summarizes the estimated VaR and TVaR at the 90%, 95%, and 98% confidence levels, comparing the empirical results with those obtained from the MSBLN-TN, Beta prime, and Trbeta distributions. The latter two were selected as the next best-fitting models based on the AIC and BIC results in Table 2. Overall, the MSBLN-TN model produces risk-measure estimates that are broadly consistent with the empirical estimates, particularly for tail-related measures. For VaR, the Trbeta distribution is closer to the empirical values at the 90% and 95% confidence levels, whereas the MSBLN-TN estimates are slightly lower. At the 98% level, MSBLN-TN provides the closest VaR estimate among the considered parametric models, while Trbeta tends to overestimate. For TVaR, MSBLN-TN is consistently closest to the empirical values across all confidence levels. By contrast, the Beta prime and Trbeta models tend to overestimate TVaR, with discrepancies becoming more pronounced in the upper tail. These results suggest that MSBLN-TN better reflects the heavy-tailed nature of the loss distribution and yields more reliable estimates of extreme risk.

Table 3. Summary of risk measures with the VaR and TVaR at 0.9, 0.95, and 0.98 security levels.

Measure	Level	Empirical	MSBLN-TN	Beta Prime	Trbeta
VaR	90%	3803.4	3677.9	3623.0	3794.8
	95%	5828.3	5476.5	5388.6	5621.1
	98%	8506.3	8774.1	8836.8	9097.1
TVaR	90%	6925.4	7297.3	7777.9	7863.7
	95%	9119.9	10,163.8	11,199.1	11,170.7
	98%	12,300.0	14,965.8	17,904.7	17,455.7

The impact of a single observation on the posterior knowledge of the latent variable λ can be seen by comparing the posterior densities (14) with $n = 1$. Specifically, based on the estimated model parameters $(\hat{\mu}, \hat{\sigma}, \hat{\mu_0}, \hat{\sigma_0})$, Figure 4 illustrates the posterior density function for each of the three representative values of x, i.e., the minimum, the median, and the maximum, given in Table 1. Clearly, the posterior of the latent variable λ becomes stochastically larger as the observed value increases.

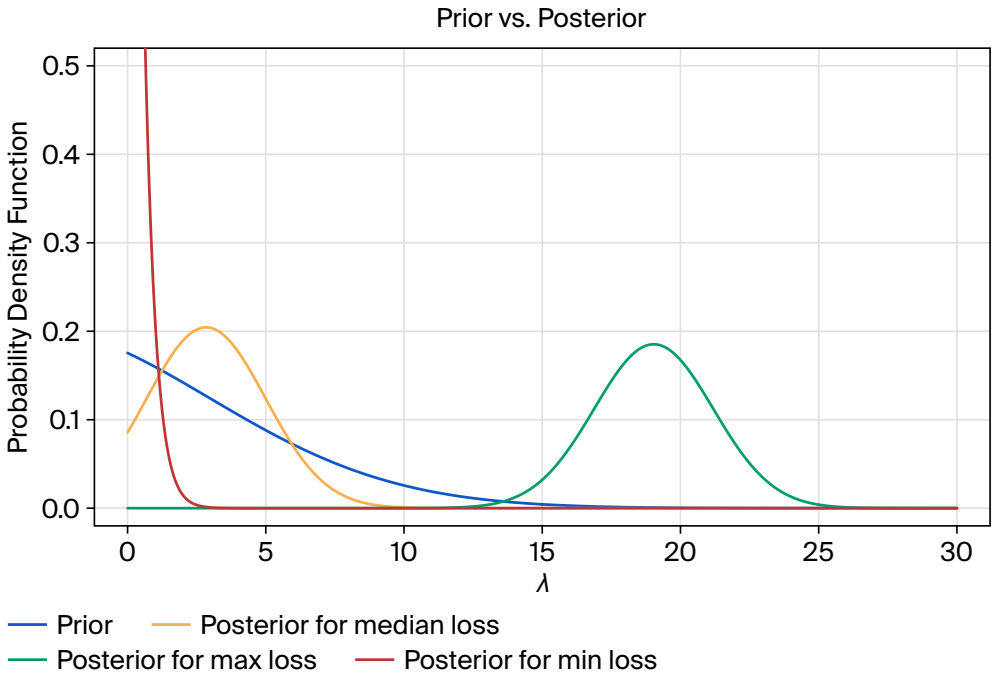


Figure 4. Prior vs. posterior on minimum, median, and maximum observations.

6.2. Regression Model Application

In this section, we apply the MSBLN-TN regression model with the EM algorithm (Algorithm A2 in Appendix D) to the `AutoClaims` dataset available in the *R* package `insuranceData`. The response variable is the amount paid to settle and close a claim (`PAID`), with $n = 6773$ observations. The input variable is the categorical risk class factor `CLASS`, with 18 levels, summarizing the operator's age, gender, marital status, and vehicle use (Frees 2009). For the initialization of the EM algorithm, all combinations of $\sigma \in \{0.5, 1\}$, $\mu_0 \in \{-1, 0, 1\}$, and $\sigma_0 \in \{0.5, 1\}$ are used, along with $\boldsymbol{\beta} = \mathbf{0}$. For each set of starting values, the iterative procedure stops when the improvement in the log-likelihood (or the stopping criterion) is less than $\epsilon = 10^{-4}$. The MSBLN-TN model parameter estimates and the corresponding log-likelihood are summarized as follows: The maximized log-likelihood is $l = -57{,}167.97$, and the estimated regression coefficients $\boldsymbol{\beta}^T = (\beta_1, \ldots, \beta_p)$ range from 3.039 to 3.705, with $\hat{\sigma} = 0.85224$, $\hat{\mu}_0 = 4.8713$, $\hat{\sigma}_0 = 0.8868$, and $\hat{\beta}_0 = 4.1087$. Since `CLASS` is a categorical variable with 18 levels, the fitted regression yields class-specific coefficient estimates $\hat{\beta}_i$. Table 4 reports these estimates along with the corresponding premiums $\hat{\mu}_{Y_i}$ for each class. The bootstrap standard errors of the estimates based on 500 replications are given in parentheses.

Level-specific importance within variable CLASS was assessed using a likelihood-ratio test (see Section 2.4). For each level k of CLASS, we fit a reduced model that collapses level k into the reference level and compared it with the full model. We quantify importance by the deviance difference $\Delta\text{Deviance} = 2(l_{full} - l_{reduced})$. These results are confirmed by Table 5. Relative to the full-level log-likelihood, risk class F7, C1C, F11, and C2 exhibit comparatively small contributions, whereas risk class C11, C71, C6, C7, C7B, and C1 show markedly large contributions and thus constitute statistically meaningful risk classes.

Figure 5 compares the fitted MSBLN-TN density curves for each risk class. Risk classes with very small contributions to the LRT, namely, F7, C1C, F11, and C2, are plotted as the red group (pick1). Those with significant contributions, namely, C11, C71, C6, C7, C7B, and C1, are plotted as the green group (pick2), and the remaining classes are plotted as the blue group (other). The red group shows a higher peak in the low value range and a faster decrease in the right tail, indicating a relatively small probability of high-value payments. Conversely, the

green group visually confirms a shift of the distribution to the right, is more dispersed, and has a heavier right tail, suggesting a higher probability of high-value payments.

Table 4. Class-specific coefficient estimates and corresponding premiums.

Risk Class (i)	C1	C11	C1A	C1B	C1C	C2
$\hat{\beta}_i$	3.4028 (0.04615)	3.4134 (0.0440)	3.3274 (0.0884)	3.4602 (0.0503)	3.2481 (0.1162)	3.2634 (0.0942)
$\hat{\mu}_{Y_i}$	1828.982 (54.980)	1848.404 (45.036)	1696.096 (136.629)	1937.074 (71.951)	1566.830 (178.949)	1590.932 (138.844)
Risk Class (i)	C6	C7	C71	C72	C7A	C7B
$\hat{\beta}_i$	3.3878 (0.0464)	3.3630 (0.0464)	3.4155 (0.0452)	3.64457 (0.0847)	3.5256 (0.0737)	3.5343 (0.0468)
$\hat{\mu}_{Y_i}$	1801.650 (46.829)	1757.663 (46.061)	1852.413 (43.916)	2329.076 (177.096)	2067.857 (138.500)	2086.044 (60.314)
Risk Class (i)	C7C	F1	F11	F6	F7	F71
$\hat{\beta}_i$	3.7054 (0.0882)	3.4658 (0.1277)	3.2659 (0.1168)	3.3723 (0.0675)	3.0389 (0.0987)	3.3973 (0.0815)
$\hat{\mu}_{Y_i}$	2475.274 (188.398)	1947.840 (246.597)	1594.995 (169.825)	1773.959 (95.214)	1271.110 (115.824)	1818.995 (131.602)

The original 18 risk classes were numerous and exhibited substantial variation in class-specific sample sizes. To address this, we clustered and merged classes with similar estimated coefficients $\hat{\beta}_i$ and implied premiums $\hat{\mu}_{Y_i}$ (see Table 4), resulting in eight risk classes. We subsequently refit the MSBLN-TN regression model using the redefined risk class as a factor to estimate the parameters. Based on these estimates, Figure 6 presents the density functions for each risk class.

Table 5. Comparison for level-specific importance within variable CLASS.

l_{full}	$l_{reduced}$					
−57,167.97	−C1	−C11	−C1A	−C1B	−C1C	−C2
	−57,564.66	−57,719.5	−57,178.06	−57,420.4	−57,173.89	−57,174.3
	−C6	−C7	−C71	−C72	−C7A	−C7B
	−57,631.06	−57,619.75	−57,713.32	−57,198.44	−57,204.19	−57,613.25
	−C7C	−F1	−F11	−F6	−F7	−F71
	−57,199.9	−57,176.6	−57,174.03	−57,211.22	−57,170.88	−57,185.74

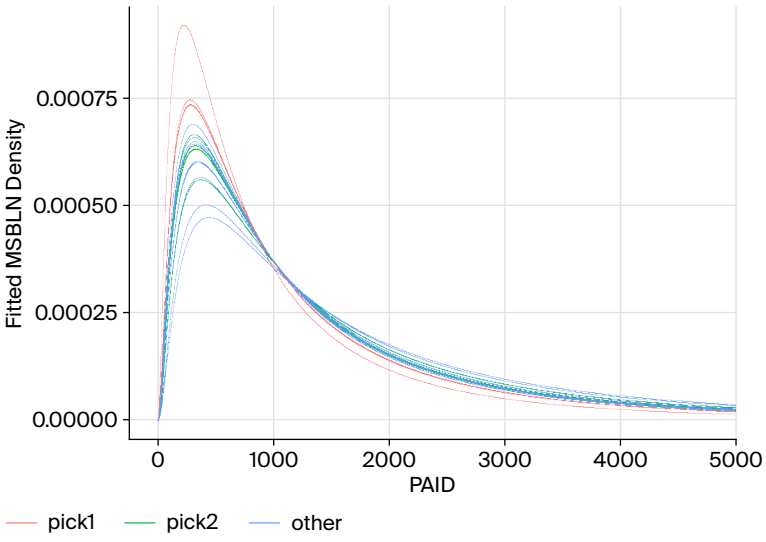


Figure 5. Risk class-specific fitted probability densities of `PAID` x: 0–5000.

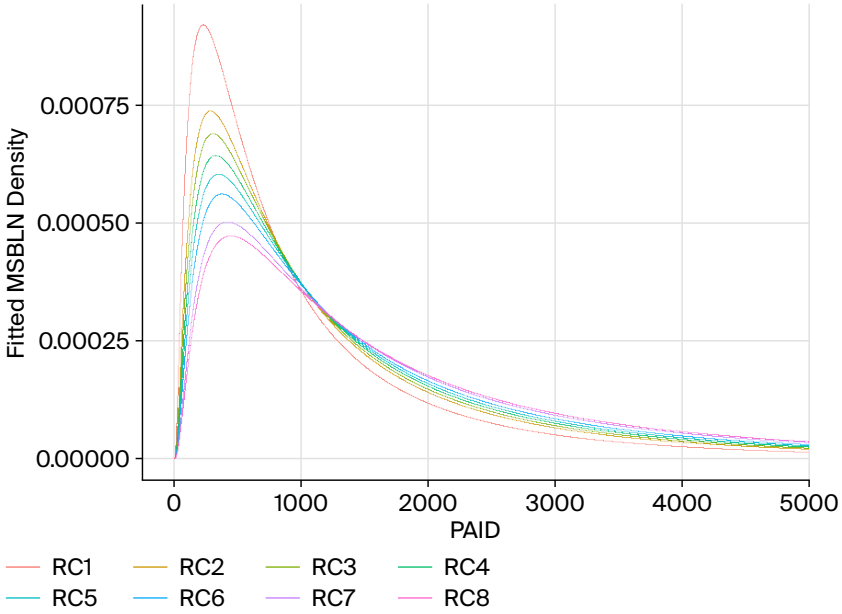


Figure 6. Risk class-specific fitted probability densities of `PAID` x: 0–5000 on merged classes.

For the redefined risk classes, Table 6 reports the risk measures VaR and TVaR, calculated using the MSBLN-TN formula with the estimated parameters. These are compared with the corresponding measures obtained without regression—that is, using only the `PAID` variable within each class. In large classes such as RC4 ($n = 5080$) and RC6 ($n = 799$), VaR and TVaR from the MSBLN-TN fit closely match their empirical counterparts. By contrast, small-sample classes exhibit substantial divergence between the two approaches. This pattern reflects the instability of empirical tail estimates under sample scarcity and indicates that the distributional assumptions of the MSBLN-TN can produce a mismatch between the model and the data in tail-risk estimation for some small-sample classes. Consequently, limited sample sizes in several redefined classes remain a data constraint that affects interpretation.

Table 6. Comparison of model-based and empirical VaR and TVaR by new risk class.

Risk Class	Level	VaR		TVaR	
		MSBLN-TN	Empirical	MSBLN-TN	Empirical
RC1 (n = 59)	90%	2825.0	2278.7	5283.5	3179.2
	95%	4164.7	2659.5	7173.9	3814.1
	98%	6446.3	3789.0	10,311.6	4073.5
RC2 (n = 139)	90%	3523.6	3092.8	6590.1	6495.3
	95%	5194.7	4557.7	8948.1	9133.3
	98%	8040.5	7409.3	12,861.8	13,242.6
RC3 (n = 77)	90%	3769.5	3890.4	70,450.0	8063.6
	95%	5557.2	6660.0	9572.4	10,457.0
	98%	8601.6	10,820.8	13,759.2	12,149.5
RC4 (n = 5080)	90%	4041.8	4088.6	7559.3	7664.8
	95%	5958.7	6300.1	10,264.0	10,304.7
	98%	9223.0	9655.1	14,753.3	14,229.7
RC5 (n = 453)	90%	4306.6	4237.5	8054.5	7688.9
	95%	6349.0	6086.9	10,936.3	10,377.1
	98%	9827.1	9181.3	15,719.7	13,741.3

Table 6. *Cont.*

Risk Class	Level	VaR		TVaR	
		MSBLN-TN	Empirical	MSBLN-TN	Empirical
RC6 (n = 799)	90%	4630.4	4866.1	8660.1	8556.4
	95%	6826.3	6681.2	11,758.6	11,299.6
	98%	10,566.0	10,121.3	16,901.6	16,162.0
RC7 (n = 85)	90%	5176.3	5492.9	9681.0	8370.8
	95%	7631.1	6213.1	13,144.8	10,486.6
	98%	11,811.6	8268.2	18,894.1	15,892.8
RC8 (n = 81)	90%	5501.2	4799.3	10,288.7	8492.1
	95%	8110.1	7273.8	13,969.9	10,852.7
	98%	12,553.1	10,123.4	20,080.1	12,679.9

7. Multivariate Extension

The proposed methodology can be readily extended to a multivariate case. For simplicity, we assume that the components of a d-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ are independent given the latent random vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)$, where $\boldsymbol{\lambda}$ has the (prior) multivariate normal distribution truncated at $\mathbf{0} = (0, \ldots, 0)$. As a result, given λ_j, X_j has an MSBLN-TN distribution with parameters μ_j and σ_j. Specifically, the conditional joint density function is given as follows:

$$\begin{aligned} g(x_1, \ldots, x_d \,|\, \lambda_1, \ldots, \lambda_d) &= \prod_{j=1}^{d} g(x_j|\lambda_j) \\ &= \frac{1}{(2\pi)^{d/2} \prod_{j=1}^{d} (x_j \sigma_j)} \exp\left\{ -\sum_{j=1}^{d} \frac{\left(\log x_j - (\mu_j + \lambda_j \sigma_j^2)\right)^2}{2\sigma_j^2} \right\}. \end{aligned}$$

The prior density of $\boldsymbol{\lambda}$ is given as

$$\pi(\lambda_1, \ldots, \lambda_d; \boldsymbol{\mu}_0, \Sigma_0) = \frac{1}{C(2\pi)^{d/2} |\Sigma_0|^{1/2}} \exp\left\{ -\frac{1}{2} (\boldsymbol{\lambda} - \boldsymbol{\mu}_0)^T \Sigma_0^{-1} (\boldsymbol{\lambda} - \boldsymbol{\mu}_0) \right\}, \boldsymbol{\lambda} > \mathbf{0}.$$

Here, the normalizing constant $C = \mathrm{P}(Y_1 > 0, \ldots, Y_d > 0)$, where $(Y_1, \ldots, Y_d)$ has the multivariate normal distribution with mean $\boldsymbol{\mu}_0$ and variance matrix Σ_0.

With these specifications, we can show that the posterior distribution of $\boldsymbol{\lambda}$ is also multivariate normal truncated at $\mathbf{0} = (0, \ldots, 0)$. Specifically, let $\boldsymbol{x}$ be the array of n random observations drawn from the proposed multivariate model. Then, the posterior density is

$$\begin{aligned} \pi(\lambda_1, \ldots, \lambda_d \,|\, \boldsymbol{x}) &\propto g(\boldsymbol{x}|\boldsymbol{\lambda}) \pi(\boldsymbol{\lambda}; \boldsymbol{\mu}_0, \Sigma_0) \\ &\propto \exp\left\{ -\frac{1}{2} (\boldsymbol{\lambda} - \boldsymbol{\mu}_0^*)^T \Sigma_0^{*-1} (\boldsymbol{\lambda} - \boldsymbol{\mu}_0^*) \right\}, \end{aligned}$$

where

$$\begin{aligned} \Sigma_0^* &= \left[\mathrm{diag}(n\sigma_1^2, \ldots, n\sigma_d^2) + \Sigma_0^{-1} \right]^{-1} \\ \boldsymbol{\mu}_0^* &= \Sigma_0^* \left[\begin{bmatrix} \sum_{i=1}^{n} (\log x_{i1} - \mu_1) \\ \vdots \\ \sum_{i=1}^{n} (\log x_{id} - \mu_d) \end{bmatrix} + \Sigma_0^{-1} \boldsymbol{\mu}_0 \right]. \end{aligned}$$

Therefore, we conclude that the multivariate truncated normal distribution is a conjugate prior of the multivariate size-biased log-normal distribution.

The unconditional joint density of $\boldsymbol{X}$ can be expressed as follows:

$$\begin{aligned} & g(x_1,\ldots,x_d \mid \boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\mu}_0,\Sigma_0) \\ = & \int_0^\infty \cdots \int_0^\infty g(x_1,\ldots,x_d \mid \lambda_1,\ldots,\lambda_d)\pi(\lambda_1,\ldots,\lambda_d;\boldsymbol{\mu}_0,\Sigma_0)d\lambda_1\cdots d\lambda_d \\ = & \frac{C^\dagger|\Sigma^\dagger|^{1/2}}{C(2\pi)^{d/2}|\Sigma_0|^{1/2}\prod_{j=1}^d x_j\sigma_j}\exp\left\{-\frac{1}{2}\left(\sum_{j=1}^d\frac{(\log x_j-\mu_j)^2}{\sigma_j^2}+\boldsymbol{\mu}_0^T\Sigma_0^{-1}\boldsymbol{\mu}_0-\boldsymbol{\mu}^{\dagger T}\Sigma^{\dagger-1}\boldsymbol{\mu}^\dagger\right)\right\}, \end{aligned}$$

where

$$\begin{aligned} \Sigma^\dagger &= \left[\mathrm{diag}(\sigma_1^2,\ldots,\sigma_d^2)+\Sigma_0^{-1}\right]^{-1} \\ \boldsymbol{\mu}^\dagger &= \Sigma^\dagger\left[\begin{bmatrix}\log x_1-\mu_1\\ \vdots\\ \log x_d-\mu_d\end{bmatrix}+\Sigma_0^{-1}\boldsymbol{\mu}_0\right] \end{aligned}$$

and $C^\dagger = \mathrm{P}(Y_1^\dagger>0,\ldots,Y_d^\dagger>0)$, where $(Y_1^\dagger,\ldots,Y_d^\dagger)$ has the multivariate normal distribution with mean $\boldsymbol{\mu}^\dagger$ and variance matrix $\Sigma^\dagger$.

Details of the derivations given in this section can be found in Appendix C.

8. Discussion

By combining the advantages of size-biased weighting, mixture models, and Bayesian ratemaking, in this study, a novel mixture distribution that can serve as an effective alternative to existing models in the context of actuarial loss modelling was proposed. To effectively capture the characteristics of extreme losses, we constructed a heavy-tailed structure using a size-biased distribution and incorporated random effects with latent variables to account for heterogeneity across individuals or units.

The resulting mixed size-biased log-normal model with a truncated normal distribution as its prior, referred to as the MSBLN-TN model, demonstrated several advantages. First, the model exhibits heavier tails than the baseline log-normal distribution, allowing for a more realistic representation of extreme losses. In addition, the tail risk measures can be expressed in a closed form in terms of the standard normal cdf, which enhances the model's practical applicability in data analysis and risk evaluation. Furthermore, the Bayesian analysis confirmed that both the prior and posterior distributions remain truncated normal, establishing a conjugate structure that improves computational efficiency and facilitates estimation process interpretability. We further extended the model to regression settings and demonstrated the applicability of the proposed methodologies with real data.

For future research, the proposed random-effects structure can be further extended to incorporate time-dependent heterogeneity, which would further improve prediction accuracy and enhance risk management efficiency in the fields of finance and insurance. The multivariate extension presented in Section 7 requires a reliable model estimation method for practical applications, and we will pursue its development in the future. In regression modelling, the risk of overfitting is of concern, especially when the number of feature variables is large. We will also consider some regularization methods to prevent overfitting and thus improve the models' predictive power.

Author Contributions: Conceptualization, T.B. and J.K.; methodology, T.B. and J.K.; software, T.B. and J.K.; validation, T.B. and J.K.; formal analysis, T.B. and J.K.; investigation, T.B. and J.K.; resources, T.B. and J.K.; data curation, J.K.; writing—original draft preparation, T.B. and J.K.; writing—review and editing, T.B., J.K. and J.Y.A.; visualization, J.K.; supervision, T.B. and J.Y.A.; project administration, T.B.; funding acquisition, T.B. and J.Y.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (RGPIN-2022-03428) (T. Bae) and the National Research Foundation of Korea (NRF) Basic Science Research Program (Priority Research Institute) (2021R1A6A1A10039823) (J. Kim and J.Y. Ahn).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset used in the severity modelling subsection is sampled from a private insurance company and thus it is not publicly available due to privacy. The dataset may be provided for a validation purpose upon request.

Acknowledgments: The authors are grateful to the anonymous referees for their helpful comments that improved the quality of the manuscript.

Conflicts of Interest: This study is a continuation of work presented in the MSc Thesis of J. Kim.

Abbreviations

The following abbreviations are used in this manuscript:

EVT	Extreme value theory
MSBLN	Mixed size-biased log-normal
MSBLN-TN	Mixed size-biased log-normal with truncated normal prior
EM	Expectation–maximization
LRT	Likelihood ratio test
CDF	Cumulative distribution function
PDF	Probability distribution function
VaR	Value at risk
TVaR	Tail value at risk
AIC	Akaike information criterion
BIC	Bayesian information criterion

Appendix A. Proof of Lemma 1

We now prove Lemma 1 by applying two variable changes and recognizing the integrand as a bivariate normal density, where Φ and ϕ denote the standard normal cdf and pdf, respectively, and $\Phi_2(\cdot,\cdot;\rho)$ denotes the CDF of a standard bivariate normal distribution with correlation ρ.

Proof. By expressing $\Phi(a+bz)$ in its integral form and interchanging the order of integration,

$$\int_c^\infty \Phi(a+bz)\,\phi(z)\,dz = \int_c^\infty \int_{-\infty}^{a+bz} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt\ \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz.$$

Next, we apply the first change of variables $z = -z_1$, so that z_1 runs from $-\infty$ to $-c$:

$$\int_c^\infty \Phi(a+bz)\,\phi(z)\,dz = \int_{-\infty}^{-c} \int_{-\infty}^{a-bz_1} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt\ \frac{1}{\sqrt{2\pi}} e^{-z_1^2/2}\,dz_1.$$

We define the second change of variables $t = \sqrt{1+b^2}\, z_2 - b z_1$ such that that for fixed z_1, the upper limit $t = a - b z_1$ corresponds to $z_2 = a/\sqrt{1+b^2}$. This yields

$$\begin{aligned}
&\int_c^{\infty} \Phi(a+bz)\, \phi(z)\, dz \\
&= \int_{-\infty}^{-c} \int_{-\infty}^{\frac{a}{\sqrt{1+b^2}}} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{(\sqrt{1+b^2} z_2 - b z_1)^2}{2} \right\} \sqrt{1+b^2} dz_2 \frac{1}{\sqrt{2\pi}} e^{-z_1^2/2} dz_1 \\
&= \int_{-\infty}^{-c} \int_{-\infty}^{\frac{a}{\sqrt{1+b^2}}} \frac{1}{2\pi\sqrt{1-\frac{b^2}{1+b^2}}} \exp\left\{ -\frac{(z_2 - \frac{b}{\sqrt{1+b^2}} z_1)^2 + (1 - \frac{b^2}{1+b^2}) z_1^2}{2(1-\frac{b^2}{1+b^2})} \right\} dz_2 dz_1.
\end{aligned}$$

Expanding the quadratic form in the exponent shows that the joint density in (z_1, z_2) is that of a bivariate normal distribution with zero means, unit variances, and correlation.

$$\rho = \frac{b}{\sqrt{1+b^2}}.$$

Therefore, the above double integral is precisely the bivariate normal cdf evaluated at the limits $z_1 \le -c$ and $z_2 \le a/\sqrt{1+b^2}$; that is,

$$\int_c^{\infty} \Phi(a+bz)\, \phi(z)\, dz = \Phi_2\left(-c,\ \frac{a}{\sqrt{1+b^2}}; \rho = \frac{b}{\sqrt{1+b^2}} \right)$$

which completes the proof of Lemma 1. □

Appendix B. Proof of Proposition 1

We now derive a closed-form expression for the Tail Value at Risk (13) under the proposed MSBLN-TN model.

Proof. By definition,

$$\mathrm{TVaR}_{\alpha}(X) = \mathbb{E}[X \mid X > \pi_{\alpha}],$$

where π_{α} denotes the α-quantile of X. Using the conditional expectation representation, we obtain

$$\begin{aligned}
\mathrm{TVaR}_{\alpha}(X) &= \mathrm{E}[X \mid X > \pi_{\alpha}] \\
&= \frac{1}{1-\alpha} \int_{\pi_{\alpha}}^{\infty} x\, g(x)\, dx \\
&= \frac{1}{1-\alpha} \int_{\pi_{\alpha}}^{\infty} x \int_0^{\infty} g(x \mid \lambda)\, \pi(\lambda)\, d\lambda\, dx \\
&= \frac{1}{1-\alpha} \int_0^{\infty} \int_{\pi_{\alpha}}^{\infty} x\, g(x \mid \lambda)\, dx\ \pi(\lambda)\, d\lambda.
\end{aligned}$$

Thus, the key step is to evaluate the inner integral $\int_{\pi_{\alpha}}^{\infty} x\, g(x \mid \lambda)\, dx$ explicitly. Since $g(x \mid \lambda)$ is the pdf of a size-biased log-normal distribution, by applying change of variable $x = e^y$, we have

$$\begin{aligned}
\int_{\pi_{\alpha}}^{\infty} x\, g(x \mid \lambda)\, dx &= \int_{\pi_{\alpha}}^{\infty} x \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{ -\frac{(\log x - \mu - \lambda\sigma^2)^2}{2\sigma^2} \right\} dx \\
&= \int_{\log \pi_{\alpha}}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{(y - \mu - \lambda\sigma^2)^2}{2\sigma^2} + y \right\} dy.
\end{aligned}$$

We now complete the square in the exponent. This yields

$$\begin{aligned}
&\int_{\log \pi_\alpha}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y-\mu-\lambda\sigma^2)^2}{2\sigma^2} + y\right\} dy \\
&= \exp\left\{\frac{(\mu+(\lambda+1)\sigma^2)^2 - (\mu+\lambda\sigma^2)^2}{2\sigma^2}\right\} \int_{\log \pi_\alpha}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y-(\mu+(\lambda+1)\sigma^2))^2}{2\sigma^2}\right\} dy \\
&= \exp\left\{\frac{2\mu+(2\lambda+1)\sigma^2}{2}\right\}\left[1-\Phi\left(\frac{\log \pi_\alpha - (\mu+(\lambda+1)\sigma^2)}{\sigma}\right)\right] \\
&= \exp\left\{\frac{2\mu+(2\lambda+1)\sigma^2}{2}\right\}\Phi\left(\frac{\mu+(\lambda+1)\sigma^2 - \log \pi_\alpha}{\sigma}\right),
\end{aligned}$$

where $\Phi(\cdot)$ denotes the standard normal cdf. Hence,

$$\int_{\pi_\alpha}^{\infty} x\, g(x \mid \lambda)\, dx = \exp\left\{\frac{2\mu+(2\lambda+1)\sigma^2}{2}\right\}\Phi\left(\frac{\mu+(\lambda+1)\sigma^2 - \log \pi_\alpha}{\sigma}\right).$$

Substituting this expression back into the TVaR integral and using the prior density $\pi(\lambda)$, we obtain

$$\begin{aligned}
\mathrm{TVaR}_\alpha(X) &= \frac{1}{1-\alpha}\int_0^{\infty}\int_{\pi_\alpha}^{\infty} x\, g(x \mid \lambda)\, dx\, \pi(\lambda)\, d\lambda \\
&= \frac{1}{1-\alpha}\int_0^{\infty} \exp\left(\mu+\lambda\sigma^2+\frac{\sigma^2}{2}\right)\Phi\left(\sigma\lambda+\sigma+\frac{\mu-\log \pi_\alpha}{\sigma}\right)\frac{\phi\big((\lambda-\mu_0)/\sigma_0\big)}{\sigma_0\,\Phi(\mu_0/\sigma_0)}\, d\lambda,
\end{aligned}$$

where $\phi(\cdot)$ is the standard normal pdf, and λ follows a normal distribution truncated to $(0,\infty)$ with mean μ_0 and variance σ_0^2.

Completing the square in the exponent with respect to λ yields

$$\begin{aligned}
&\mathrm{TVaR}_\alpha(X) \\
&= \frac{\exp\left(\mu+\frac{\sigma^2}{2}+\mu_0\sigma^2+\frac{\sigma_0^2\sigma^4}{2}\right)}{(1-\alpha)\sigma_0\,\Phi(\frac{\mu_0}{\sigma_0})}\int_0^{\infty}\Phi\left(\sigma\lambda+\sigma+\frac{\mu-\log \pi_\alpha}{\sigma}\right)\phi\left(\frac{\lambda-(\mu_0+\sigma_0^2\sigma^2)}{\sigma_0}\right)d\lambda.
\end{aligned}$$

For the integral in the above expression, we standardize the mixing variable by setting

$$z = \frac{\lambda-(\mu_0+\sigma_0^2\sigma^2)}{\sigma_0},$$

such that $\lambda = \mu_0+\sigma_0^2\sigma^2+\sigma_0 z$ and $\lambda \in (0,\infty)$ corresponds to $z \in \big(-(\mu_0+\sigma_0^2\sigma^2)/\sigma_0, \infty\big)$. This leads to

$$\begin{aligned}
&\int_0^{\infty}\Phi\left(\sigma\lambda+\sigma+\frac{\mu-\log \pi_\alpha}{\sigma}\right)\phi\left(\frac{\lambda-(\mu_0+\sigma_0^2\sigma^2)}{\sigma_0}\right)d\lambda \\
&= \int_{-\frac{\mu_0+\sigma_0^2\sigma^2}{\sigma_0}}^{\infty}\Phi\left(\sigma_0\sigma z+\sigma(\mu_0+\sigma_0^2\sigma^2+1)+\frac{\mu-\log \pi_\alpha}{\sigma}\right)\phi(z)\, dz.
\end{aligned}$$

The remaining integral is of the form $\int_c^{\infty}\Phi(a+bz)\,\phi(z)\,dz$ and can be evaluated using Lemma 1. Hence,

$$\mathrm{TVaR}_\alpha(X) = \frac{\exp\left(\mu+\frac{\sigma^2}{2}+\mu_0\sigma^2+\frac{\sigma_0^2\sigma^4}{2}\right)\Phi_2\Big(-c,\ \frac{a}{\sqrt{1+b^2}};\ \rho=\frac{b}{\sqrt{1+b^2}}\Big)}{(1-\alpha)\,\Phi\left(\frac{\mu_0}{\sigma_0}\right)},$$

where

$$a = \sigma(\mu_0+\sigma_0^2\sigma^2+1)+\frac{\mu-\log \pi_\alpha}{\sigma}, \qquad b = \sigma_0\sigma, \qquad c = -\frac{\mu_0+\sigma_0^2\sigma^2}{\sigma_0},$$

and $\Phi_2(\cdot,\cdot;\rho)$ is the cdf of a standard bivariate normal distribution with correlation $\rho = b/\sqrt{1+b^2}$. □

Appendix C. Derivations for Section 7

Appendix C.1. Posterior Density

The posterior density of $\boldsymbol{\lambda}$ is proportional to the product of the likelihood and the prior. Collecting terms in $\boldsymbol{\lambda}$ and completing the square yield a multivariate truncated normal posterior distribution.

$$
\begin{aligned}
&\pi(\lambda_1,\ldots,\lambda_d \,|\, \boldsymbol{x}) \\
&\propto g(\boldsymbol{x}|\boldsymbol{\lambda})\pi(\boldsymbol{\lambda};\boldsymbol{\mu}_0,\Sigma_0) \\
&\propto \exp\left\{-\sum_{i=1}^{n}\sum_{j=1}^{d}\frac{\left(\log x_{ij}-(\mu_j+\lambda_j\sigma_j^2)\right)^2}{2\sigma_j^2}\right\}\exp\left\{-\frac{1}{2}(\boldsymbol{\lambda}-\boldsymbol{\mu}_0)^T\Sigma_0^{-1}(\boldsymbol{\lambda}-\boldsymbol{\mu}_0)\right\} \\
&\propto \exp\left\{-\sum_{j=1}^{d}\frac{\sigma_j^2}{2}\sum_{i=1}^{n}\left(\lambda_j-\frac{\log x_{ij}-\mu_j}{\sigma_j^2}\right)^2\right\}\exp\left\{-\frac{1}{2}(\boldsymbol{\lambda}-\boldsymbol{\mu}_0)^T\Sigma_0^{-1}(\boldsymbol{\lambda}-\boldsymbol{\mu}_0)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\sum_{j=1}^{d}\left[n\sigma_j^2\lambda_j^2-2\left(\sum_{i=1}^{n}(\log x_{ij}-\mu_j)\right)\lambda_j\right]\right\}\exp\left\{-\frac{1}{2}(\boldsymbol{\lambda}-\boldsymbol{\mu}_0)^T\Sigma_0^{-1}(\boldsymbol{\lambda}-\boldsymbol{\mu}_0)\right\} \\
&\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\lambda}-\boldsymbol{\mu}_1)^T\Sigma_1^{-1}(\boldsymbol{\lambda}-\boldsymbol{\mu}_1)\right\}\exp\left\{-\frac{1}{2}(\boldsymbol{\lambda}-\boldsymbol{\mu}_0)^T\Sigma_0^{-1}(\boldsymbol{\lambda}-\boldsymbol{\mu}_0)\right\} \\
&\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\lambda}-\boldsymbol{\mu}_0^*)^T\Sigma_0^{*-1}(\boldsymbol{\lambda}-\boldsymbol{\mu}_0^*)\right\},
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{\mu}_1 &= \begin{bmatrix}\sum_{i=1}^{n}(\log x_{i1}-\mu_1)/n\sigma_1^2 \\ \vdots \\ \sum_{i=1}^{n}(\log x_{id}-\mu_d)/n\sigma_d^2\end{bmatrix}, \\
\Sigma_1^{-1} &= \operatorname{diag}(n\sigma_1^2,\ldots,n\sigma_d^2), \\
\Sigma_0^{*-1} &= \Sigma_1^{-1}+\Sigma_0^{-1} = \operatorname{diag}(n\sigma_1^2,\ldots,n\sigma_d^2)+\Sigma_0^{-1}, \\
\boldsymbol{\mu}_0^* &= \Sigma_0^*\left(\Sigma_1^{-1}\boldsymbol{\mu}_1+\Sigma_0^{-1}\boldsymbol{\mu}_0\right) = \Sigma_0^*\begin{bmatrix}\sum_{i=1}^{n}(\log x_{i1}-\mu_1) \\ \vdots \\ \sum_{i=1}^{n}(\log x_{id}-\mu_d)\end{bmatrix}+\Sigma_0^*\Sigma_0^{-1}\boldsymbol{\mu}_0.
\end{aligned}
$$

Appendix C.2. Unconditional Density

We derive the unconditional joint density by integrating out the latent vector $\boldsymbol{\lambda} = (\lambda_1,\ldots,\lambda_d)$. After completing the square, the remaining integral corresponds to a normalized truncated multivariate Gaussian density and thus equals 1, yielding the expression below.

$$
\begin{aligned}
& g(\boldsymbol{x} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\mu}_0, \Sigma_0) \\
&= \int_0^\infty \cdots \int_0^\infty g(\boldsymbol{x} \mid \boldsymbol{\lambda}) \pi(\boldsymbol{\lambda}; \boldsymbol{\mu}_0, \Sigma_0) d\boldsymbol{\lambda} \\
&= \int_0^\infty \cdots \int_0^\infty \frac{\exp\left\{ -\sum_{j=1}^d \frac{\left(\log x_j - (\mu_j + \lambda_j \sigma_j^2)\right)^2}{2\sigma_j^2} \right\}}{(2\pi)^{d/2} \prod_{j=1}^d (x_j \sigma_j)} \frac{\exp\left\{ -\frac{1}{2} (\boldsymbol{\lambda} - \boldsymbol{\mu}_0)^T \Sigma_0^{-1} (\boldsymbol{\lambda} - \boldsymbol{\mu}_0) \right\}}{C(2\pi)^{d/2} |\Sigma_0|^{1/2}} d\boldsymbol{\lambda} \\
&= \frac{1}{C|\Sigma_0|^{1/2}(2\pi)^d/2 \prod_{j=1}^d (x_j \sigma_j)} \exp\left\{ -\frac{1}{2} \left(\sum_{j=1}^d \frac{(\log x_j - \mu_j)^2}{\sigma_j^2} + \boldsymbol{\mu}_0^T \Sigma_0^{-1} \boldsymbol{\mu}_0 \right) \right\} \\
&\times \int_0^\infty \cdots \int_0^\infty \exp\left\{ -\frac{1}{2} \left[\sum_{j=1}^d \left(\sigma_j^2 \lambda_j^2 - 2(\log x_j - \mu_j) \lambda_j \right) + \left(\boldsymbol{\lambda}^T \Sigma_0^{-1} \boldsymbol{\lambda} - 2 \boldsymbol{\lambda}^T \Sigma_0^{-1} \boldsymbol{\mu}_0 \right) \right] \right\} d\boldsymbol{\lambda} \\
&= \frac{1}{C|\Sigma_0|^{1/2}(2\pi)^d/2 \prod_{j=1}^d (x_j \sigma_j)} \exp\left\{ -\frac{1}{2} \left(\sum_{j=1}^d \frac{(\log x_j - \mu_j)^2}{\sigma_j^2} + \boldsymbol{\mu}_0^T \Sigma_0^{-1} \boldsymbol{\mu}_0 \right) \right\} \\
&\times \int_0^\infty \cdots \int_0^\infty \exp\left\{ -\frac{1}{2} \left[\boldsymbol{\lambda}^T \left(\mathrm{diag}(\sigma_1^2, \ldots, \sigma_d^2) + \Sigma_0^{-1} \right) \boldsymbol{\lambda} - 2 \boldsymbol{\lambda}^T \left((\log \boldsymbol{x} - \boldsymbol{\mu}) + \Sigma_0^{-1} \boldsymbol{\mu}_0 \right) \right] \right\} d\boldsymbol{\lambda} \\
&= \frac{C^\dagger |\Sigma^\dagger|^{1/2}}{C(2\pi)^{d/2} |\Sigma_0|^{1/2} \prod_{j=1}^d x_j \sigma_j} \exp\left\{ -\frac{1}{2} \left(\sum_{j=1}^d \frac{(\log x_j - \mu_j)^2}{\sigma_j^2} + \boldsymbol{\mu}_0^T \Sigma_0^{-1} \boldsymbol{\mu}_0 - \boldsymbol{\mu}^{\dagger T} \Sigma^{\dagger -1} \boldsymbol{\mu}^\dagger \right) \right\} \\
&\times \int_0^\infty \cdots \int_0^\infty \frac{\exp\left\{ -\frac{1}{2} (\boldsymbol{\lambda} - \boldsymbol{\mu}^\dagger)^T \Sigma^{\dagger -1} (\boldsymbol{\lambda} - \boldsymbol{\mu}^\dagger) \right\}}{C^\dagger (2\pi)^{d/2} |\Sigma^\dagger|^{1/2}} d\boldsymbol{\lambda} \\
&= \frac{C^\dagger |\Sigma^\dagger|^{1/2}}{C(2\pi)^{d/2} |\Sigma_0|^{1/2} \prod_{j=1}^d x_j \sigma_j} \exp\left\{ -\frac{1}{2} \left(\sum_{j=1}^d \frac{(\log x_j - \mu_j)^2}{\sigma_j^2} + \boldsymbol{\mu}_0^T \Sigma_0^{-1} \boldsymbol{\mu}_0 - \boldsymbol{\mu}^{\dagger T} \Sigma^{\dagger -1} \boldsymbol{\mu}^\dagger \right) \right\}
\end{aligned}
$$

where

$$
\begin{aligned}
\Sigma^\dagger &= \left[\mathrm{diag}(\sigma_1^2, \ldots, \sigma_d^2) + \Sigma_0^{-1} \right]^{-1}, \\
\boldsymbol{\mu}^\dagger &= \Sigma^\dagger \begin{bmatrix} \log x_1 - \mu_1 \\ \vdots \\ \log x_d - \mu_d \end{bmatrix} + \Sigma^\dagger \Sigma_0^{-1} \boldsymbol{\mu}_0.
\end{aligned}
$$

Appendix D. Estimation Algorithms

Algorithm A1 MSBLN-TN model fitting algorithm

1 Prepare K sets of initial parameters $\{\boldsymbol{\theta}_k^{(0)}\}_{k=1}^K$

2 **for** $k = 1, \ldots, K$ **do**

while the stopping criterion $D > \epsilon$ **do**

Obtain the updated estimates $\boldsymbol{\theta}^{(s+1)} = (\mu^{(s+1)}, \sigma^{(s+1)}, \mu_0^{(s+1)}, \sigma_0^{(s+1)})$

$$\begin{aligned}(\sigma^{(s+1)})^2 &= \frac{1}{2(\bar{b}-\bar{a}^2)}\left(-1+\sqrt{1+4(\bar{b}-\bar{a}^2)\left(n^{-1}\sum_{i=1}^{n}(\log x_i)^2-\bar{l}^2\right)}\right)\\ \mu^{(s+1)} &= \bar{l}-\bar{a}(\sigma^{(s+1)})^2,\end{aligned}$$

where $\bar{l} = n^{-1}\sum_{i=1}^{n}\log x_i$,

$$\begin{aligned}\bar{a} &= \frac{1}{n}\sum_{i=1}^{n}\mu_{0i}^{*(s)}+\frac{\sigma_0^{*(s)}}{n}\sum_{i=1}^{n}\frac{\phi(\mu_{0i}^{*(s)}/\sigma_0^{*(s)})}{\Phi(\mu_{0i}^{*(s)}/\sigma_0^{*(s)})},\\ \bar{b} &= (\sigma_0^{*(s)})^2+\frac{1}{n}\sum_{i=1}^{n}(\mu_{0i}^{*(s)})^2+\frac{\sigma_0^{*(s)}}{n}\sum_{i=1}^{n}\mu_{0i}^{*(s)}\frac{\phi(\mu_{0i}^{*(s)}/\sigma_0^{*(s)})}{\Phi(\mu_{0i}^{*(s)}/\sigma_0^{*(s)})},\end{aligned}$$

and

$$\begin{aligned}\mu_{0i}^{*(s)} &= \frac{\mu_0^{(s)}-\mu^{(s)}(\sigma_0^{(s)})^2+(\sigma_0^{(s)})^2\log x_i}{1+(\sigma^{(s)})^2(\sigma_0^{(s)})^2}, i=1,\ldots,n,\\ \sigma_0^{*(s)} &= \frac{\sigma_0^{(s)}}{\sqrt{1+(\sigma^{(s)})^2(\sigma_0^{(s)})^2}}.\end{aligned}$$

$(\mu_0^{(s+1)}, \sigma_0^{(s+1)})$ is the solution of the system of equations:

$$\begin{cases}\mu_0+\sigma_0\frac{\phi(\mu_0/\sigma_0)}{\Phi(\mu_0/\sigma_0)}=\bar{a}\\ \bar{a}\mu_0+\sigma_0^2=\bar{b},\end{cases}$$

with the constraint $\mu_0 < \bar{b}/\bar{a}$.

Compute the log-likelihood function at updated estimates:

$l^{(s+1)} = \sum_{i=1}^{n}\log g(x_i; \mu^{(s+1)}, \sigma^{(s+1)}, \mu_0^{(s+1)}, \sigma_0^{(s+1)})$.

Calculate the criterion $D \leftarrow (l^{(s+1)} - l^{(s)})$

end

Record the outputs: $\text{Loglike}_k = l^{(s+1)}$ and $\hat{\boldsymbol{\theta}}_k = \boldsymbol{\theta}^{(s+1)}$.

end

3 Return the final fitted model corresponding to $\max_{k=1,\ldots,K}\text{Loglike}_k$.

Algorithm A2 MSBLN-TN regression model fitting algorithm

1 Prepare K sets of initial parameters $\{\boldsymbol{\theta}_k^{(0)} = (\sigma^{(0)}, \mu_0^{(0)}, \sigma_0^{(0)})\}_{k=1}^K$ and $\boldsymbol{\beta}^{(0)}$

2 **for** $k = 1, \ldots, K$ **do**

while the stopping criterion $D > \epsilon$ **do**

Obtain the updated estimates $\boldsymbol{\beta}^{(s+1)}$ and $\boldsymbol{\theta}^{(s+1)} = (\sigma^{(s+1)}, \mu_0^{(s+1)}, \sigma_0^{(s+1)})$:

$$
\begin{aligned}
(\sigma^{(s+1)})^2 &= \frac{1}{2(\bar{b}-\bar{f})}\left(-1+\sqrt{1+4(\bar{b}-\bar{f})n^{-1}\sum_{i=1}^{n}(\log y_i - d_i)^2}\right)\\
\boldsymbol{\beta}^{(s+1)} &= \left(\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^T\right)^{-1}\left(\boldsymbol{\zeta}-\sigma^2\boldsymbol{\eta}\right),
\end{aligned}
$$

where

$$
\begin{aligned}
\bar{b} &= (\sigma_0^{*(s)})^2 + \frac{1}{n}\sum_{i=1}^{n}(\mu_{0i}^{*(s)})^2 + \frac{\sigma_0^{*(s)}}{n}\sum_{i=1}^{n}\mu_{0i}^{*(s)}\frac{\phi(\mu_{0i}^{*(s)}/\sigma_0^{*(s)})}{\Phi(\mu_{0i}^{*(s)}/\sigma_0^{*(s)})},\\
\bar{f} &= n^{-1}\sum_{i=1}^{n}e_i^2,\\
e_i &= \boldsymbol{x}_i^T\left(\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^T\right)^{-1}\boldsymbol{\eta}, \quad i=1,\ldots,n,\\
\boldsymbol{\eta} &= \sum_{i=1}^{n}\left(\mu_{0i}^{*(s)}+\sigma_0^{*(s)}\frac{\phi(\mu_{0i}^{*(s)}/\sigma_0^{*(s)})}{\Phi(\mu_{0i}^{*(s)}/\sigma_0^{*(s)})}\right)\boldsymbol{x}_i,\\
d_i &= \boldsymbol{x}_i^T\left(\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^T\right)^{-1}\boldsymbol{\zeta}, \quad i=1,\ldots,n,\\
\boldsymbol{\zeta} &= \sum_{i=1}^{n}(\log y_i)\boldsymbol{x}_i,\\
\mu_{0i}^{*(s)} &= \frac{\mu_0^{(s)} - \boldsymbol{x}_i^T\boldsymbol{\beta}^{(s)}(\sigma_0^{(s)})^2 + (\sigma_0^{(s)})^2\log y_i}{1+(\sigma^{(s)})^2(\sigma_0^{(s)})^2}, \; i=1,\ldots,n,\\
\sigma_0^{*(s)} &= \frac{\sigma_0^{(s)}}{\sqrt{1+(\sigma^{(s)})^2(\sigma_0^{(s)})^2}}.
\end{aligned}
$$

$(\mu_0^{(s+1)}, \sigma_0^{(s+1)})$ is the solution of the system of equations:

$$
\begin{cases}
\mu_0 + \sigma_0\frac{\phi(\mu_0/\sigma_0)}{\Phi(\mu_0/\sigma_0)} = \bar{a}\\
\bar{a}\mu_0 + \sigma_0^2 = \bar{b},
\end{cases}
$$

with the constraint $\mu_0 < \bar{b}/\bar{a}$, where
$\bar{a} = \frac{1}{n}\sum_{i=1}^{n}\mu_{0i}^{*(s)} + \frac{\sigma_0^{*(s)}}{n}\sum_{i=1}^{n}\frac{\phi(\mu_{0i}^{*(s)}/\sigma_0^{*(s)})}{\Phi(\mu_{0i}^{*(s)}/\sigma_0^{*(s)})}$.
Compute the log-likelihood function at the updated estimates:
$l^{(s+1)} = \sum_{i=1}^{n}\log g(y_i\,;\, x_i^T\boldsymbol{\beta}^{(s+1)}, \sigma^{(s+1)}, \mu_0^{(s+1)}, \sigma_0^{(s+1)})$.
Calculate the criterion $D \leftarrow (l^{(s+1)} - l^{(s)})$
end

Record the outputs: $\text{Loglike}_k = l^{(s+1)}$, $\hat{\boldsymbol{\beta}}_k = \boldsymbol{\beta}^{(s+1)}$ and $\hat{\boldsymbol{\theta}}_k = \boldsymbol{\theta}^{(s+1)}$.

end

3 Return the final fitted model corresponding to $\max_{k=1,\ldots,K}\text{Loglike}_k$.

References

Bae, Taehan, and Bangwon Ko. 2020. On the mixtures of length-biased Weibull distributions for loss severity modeling. *Journal of the Korean Statistical Society* 49: 422–38. [CrossRef]

Bae, Taehan, and Tatjana Miljkovic. 2024. Loss modeling with the size-biased lognormal mixture and the entropy regularized EM algorithm. *Insurance: Mathematics and Economics* 117: 182–95. [CrossRef]

Blostein, Martin, and Tatjana Miljkovic. 2019. On modeling left-truncated loss data using mixtures of distributions. *Insurance: Mathematics and Economics* 85: 35–46. [CrossRef]

De Jong, Piet, and Gillian Z. Heller. 2008. *Generalized Linear Models for Insurance Data*. Cambridge : Cambridge University Press.

Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* 39: 1–22. [CrossRef]

Embrechts, Paul, Claudia Klüppelberg, and Thomas Mikosch. 1997. *Modelling Extremal Events for Insurance and Finance.* Berlin: Springer.

Foss, Sergey, Dmitry Korshunov, and Stan Zachary. 2011. *An Introduction to Heavy-Tailed and Subexponential Distributions.* Berlin and Heidelberg: Springer.

Frees, Edward W. 2009. *Regression Modeling with Actuarial and Financial Applications*. Cambridge: Cambridge University Press.

Frees, Edward W., and Emiliano A. Valdez. 1998. Understanding relationships using copulas. *The North American Actuarial Journal* 2: 1–25. [CrossRef]

Gupta, Ramesh C., and S. N. U. A. Kirmani. 1990. The role of weighted distributions in stochastic modeling. *Communications in Statistics-Theory and Methods* 19: 3147–62. [CrossRef]

Hogg, Robert V., and Stuart A. Klugman . 1984. *Loss Distributions*. Hoboken: John Wiley & Sons.

Klugman, Stuart A., Harry H. Panjer, and Gordon E. Willmot. 2019. *Loss Models: From Data to Decisions*, 5th ed. New York: John Wiley & Sons.

Lee, Simon C. K., and Sheldon Lin. 2010. Modeling and evaluating insurance losses via mixtures of Erlang distributions. *North American Actuarial Journal* 14: 107–30. [CrossRef]

Patil, Ganapati P., and Calyampudi R. Rao. 1978. Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics* 34: 179–89. [CrossRef]

Reynkens, Tom, Roel Verbelen, Jan Beirlant, and Katrien Antonio. 2017. Modelling censored losses using splicing: A global fit strategy with mixed Erlang and extreme value distributions. *Insurance: Mathematics and Economics* 77: 65–77. [CrossRef]

Tseung, Spark C., Ian Weng Chan, Tsz Chai Fung, Andrei L. Badescu, and Sheldon Lin. 2023. Improving risk classification and ratemaking using mixture-of-experts models with random effects. *Journal of Risk and Insurance* 90: 789–820. [CrossRef]

Verbelen, Roel, Lan Gong, Katrien Antonio, Andrei Badescu, and Sheldon Lin. 2015. Fitting mixtures of Erlangs to censored and truncated data using the EM algorithm. *ASTIN Bulletin: The Journal of the IAA* 45: 729–58. [CrossRef]

Willmot, Gordon E., and Sheldon Lin. 2011. Risk modelling with the mixed Erlang distribution. *Applied Stochastic Models in Business and Industry* 27: 2–16. [CrossRef]

Yin, Cuihong, and Sheldon Lin. 2016. Efficient estimation of Erlang mixtures using iSCAD penalty with insurance application. *ASTIN Bulletin: The Journal of the IAA* 46: 779–99. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.