Research Article

On the Usefulness of the Logarithmic Skew Normal Distribution for Describing Claims Size Data

E. Gómez–Déniz [1] **and E. Calderín-Ojeda** [2]

^{1}*Department of Quantitative Methods in Economics and Tides Institute, University of Las Palmas de Gran Canaria,*
Las Palmas de Gran Canaria 35017, Spain
^{2}*Centre for Actuarial Studies, Department of Economics, The University of Melbourne, Melbourne, Australia*

Correspondence should be addressed to E. Calderín-Ojeda; ecalderin@unimelb.edu.au

In this paper, the three-parameter skew lognormal distribution is proposed to model actuarial data concerning losses. This distribution yields a satisfactory fit to empirical data in the whole range of the empirical distribution as compared to other distributions used in the actuarial statistics literature. To the best of our knowledge, this distribution has not been used in insurance context and it might be suitable for computing reinsurance premiums in situations where the right tail of the empirical distribution plays an important role. Furthermore, a regression model can be simply derived to explain the response variable as a function of a set of explanatory variables.

1. Introduction

The modelling of large claims is a topic of relevant importance in general insurance and reinsurance, particularly in the field of mathematical risk theory, see, for instance, McNeil [1]; Beirlant and Teugels [2]; Beirlant et al. [3]; and Embrechts et al. [4]. For a comprehensive study about reinsurance, see the recent book of Albrecher et al. [5]. It is our interest to find simple statistical distributions appropriate for modelling both, smaller and medium-size losses with a high frequency and large losses with a low frequency. This is an issue of particular interest in the context of reinsurance and premium calculation principles. In this sense, the classical Pareto distribution [6–8] has been traditionally considered as a suitable claims' size distribution in relation to rating problems. Concerning this, the single parameter Pareto distribution not only has nice statistical properties but also provides a good description of the random behaviour of large losses (e.g., the right tail of the distribution). Particularly, when calculating deductibles and excess-of-loss levels of reinsurance, the simple Pareto distribution has been proved to be convenient (see, for example, [9, 10]). As an alternative to the classical Pareto distribution, other models

have been recently introduced in the actuarial literature by Sarabia et al. [11–13] and Ghitany et al. [14].

On the contrary, there exist many situations where the empirical data show slight or marked asymmetry. This is frequently the case, for example, with actuarial and financial data that also have heavy tails reflecting the existence of extreme values. These two features imply that the data cannot be adequately modelled by the Gaussian or normal distribution.

Let g and G, respectively, be the probability density function (pdf) and the cumulative distribution function (cdf) of a symmetric distribution. A random variable Z is said to have a skew distribution if its pdf is given by

$$f_Z(z) = 2g(z)G(\gamma z), \quad -\infty < z < \infty, \ \gamma \in \mathbb{R}. \quad (1)$$

This family of distributions has been widely studied as an extension of the normal distribution via a shape parameter, γ, that accounts for the skewness of the model. When $g(\cdot)$ and $G(\cdot)$ are replaced in (1) by $\phi(z)$ and $\Phi(z)$, i.e., the standard normal density and distribution function, respectively, the resulting model is called the skew normal distribution.

In this paper, special attention is paid to the generalized skew normal density provided in Henze [15] and also studied by Arnold and Beaver [16]. Its pdf is given by

$$f_Z(z) = \frac{\phi(z)\Phi(\gamma_0 + \gamma_1 z)}{\Phi\left(\gamma_0 / \left(\sqrt{1 + \gamma_1^2}\right)\right)}, \quad -\infty < z < \infty, \ \gamma_0, \gamma_1 \in \mathbb{R}. \tag{2}$$

For multivariate extensions, see for instance Azzalini and Valle [17], Azzalini and Capitanio [18], and Arnold and Beaver [16]. For an exhaustive and comprehensive study of the skew-normal distribution, see the recent book of Azzalini [19].

This paper is organized as follows. In Section 2, the proposed distribution is studied and some interesting properties are given. Numerical applications are provided in Section 3, and conclusions are drawn in Section 4.

2. Modelling the Size of Losses

The basic skew lognormal distribution has been studied by Lin and Stoyanov [20] (see also [19]; Chap. 2, [21–23]). Nevertheless, in this paper, we will pay special attention to the distribution arising from exponentiation of (2) in the following sense: if X is a random variable with density (2), $X \sim f_Z$, we consider a new random variable $Y \sim \exp(X)$. Also, if taking $\gamma_0 = \gamma_1 = \lambda$ and using a linear transformation, we allow for more flexible location and scale parameters in our model. Thus, we get the pdf given by

$$f_{\mu,\sigma,\lambda}(x) = \frac{\phi(r_{\mu,\sigma}(x))}{\sigma x} \frac{\Phi((1 + r_{\mu,\sigma}(x))\lambda)}{\Phi(\lambda_0)}, \quad x > 0, \tag{3}$$

where

$$r_{\mu,\sigma}(x) = \frac{\log x - \mu}{\sigma}, \tag{4}$$
$$\lambda_0 = \frac{\lambda}{\sqrt{1 + \lambda^2}}.$$

Here, $\lambda \in \mathbb{R}$, $\mu \in \mathbb{R}$, and $\sigma > 0$. Observe that when $\lambda = 0$ expression (3) reduces to the classical lognormal distribution. It can be seen that the parameter λ regulates the shape of the distribution. Furthermore, Lin and Stoyanov [20] established that the distribution has heavy tails and therefore it is a suitable distribution to be incorporated to the wide catalogue of heavy tails (see [5]). Additionally Lin and Stoyanov [20] provided a stochastic representation of this distribution and determined that the distribution can be obtained as the product of two independent random variables, one of which is lognormal and the other one a logarithmic half-normal. Simple calculations show that density (3) can be also written as

$$f_{\mu,\sigma,\lambda}(x) = \frac{\exp\{1/2(1 + 2r_{\mu,\sigma}(x))\}}{\Phi(\lambda_0)} f_Z(z), \tag{5}$$

where $f_Z(z)$ is expression (1) with $z = (1 + r_{\mu,\sigma}(x))$, $\gamma = \lambda$, $g(z) = \phi(z)$, and $G(z) = \Phi(z)$.

It is straightforward to show that the distribution is unimodal with the mode satisfying the equation:

$$\frac{\lambda \phi(1 + r_{\mu,\sigma}(z))}{\Phi(1 + r_{\mu,\sigma}(z))} = r_{\mu,\sigma}(z) + \sigma. \tag{6}$$

The mean value and the second order moment of the random variable X with density (3), $X \sim \mathrm{LSN}(\mu, \sigma, \lambda)$, are given by

$$E[X] = \frac{\Phi((1 + \sigma)\lambda_0)}{\Phi(\lambda_0)} \exp\left\{\mu + \frac{\sigma^2}{2}\right\}, \tag{7}$$
$$E[X^2] = \frac{\Phi((1 + 2\sigma)\lambda_0)}{\Phi(\lambda_0)} \exp\{2(\mu + \sigma^2)\}.$$

Let $F_{\mu,\sigma,\lambda}(x) = \Pr[X \le x]$ denote the cdf of $X \sim \mathrm{LSN}(\mu, \sigma, \lambda)$. By applying directly result B.21 in Azzalini [19], it is not difficult to observe that this cdf is given by

$$F_{\mu,\sigma,\lambda}(x) = \Phi(r_{\mu,\sigma}(x)) + \frac{1}{\Phi(\lambda_0)}\left[T\left(r_{\mu,\sigma}(x), \frac{\lambda_0}{r_{\mu,\sigma}(x)}\right) + T\left(\lambda_0, \frac{r_{\mu,\sigma}(x)}{\lambda_0}\right) \right.$$
$$\left. - T\left(r_{\mu,\sigma}(x), \frac{(1 + r_{\mu,\sigma}(x))\lambda}{r_{\mu,\sigma}(x)}\right) - T\left(\lambda_0, \frac{\lambda^2 + (1 + \lambda^2)r_{\mu,\sigma}(x)}{\lambda}\right) \right], \tag{8}$$

which is satisfied whenever $\lambda \ne 0$ and $x > 0$. Here, $T(x, a)$ represents Owen's function (see [24]) given by

$$T(x, a) = \frac{1}{2\pi} \int_0^a \frac{1}{1 + t^2} \exp\left\{-\frac{1}{2}x^2(1 + t^2)\right\} dt, \quad a \in \mathbb{R}^+. \tag{9}$$

If $\lambda = 0$, we deal with the standard normal distribution, and $T(x, 0) = 0$. The latter function can be easily computed by using the Mathematica software package.

The Acceptance-Rejection method of simulation can be used to generate random variates from (3). We begin by simulating a value from a lognormal distribution with parameters μ and σ. Then, having chosen an alternative random variable that follows a lognormal probability distribution to simulate from, we define a constant c in the following way:

$$c = \max_x \frac{f_{\mu,\sigma,\lambda}(x)}{g(x)}, \tag{10}$$

where $f_{\mu,\sigma,\lambda}(x)$ is given by (3) and $g(x)$ is the standard lognormal density. The algorithm for simulating a random variate from the LSN distribution is as follows:

(1) Generate a random variate from the lognormal distribution, x_1.

(2) Generate a random number u_1.

(3) If $u_1 \leq \Phi((1 + r_{\mu,\sigma}(x_1))\lambda)$, then set the simulated value from (3) equal to x_1. Otherwise, return to step 1.

By writing μ as

$$\mu = -\frac{\sigma^2}{2} + \log\left[\frac{\theta\,\Phi(\lambda_0)}{\Phi((1+\sigma)\lambda_0)}\right], \quad (11)$$

pdf (3) has mean value equal to $\theta > 0$ and variance provided by

$$\mathrm{var}(X) = \psi_1(\theta)\psi_2(\sigma, \lambda), \quad (12)$$

where

$$\psi_1(\theta) = \theta^2,$$

$$\psi_2(\sigma, \lambda) = \frac{\Phi((1+2\sigma)\lambda_0)[\Phi(\lambda_0)]^2\exp\{-\sigma^2\}}{[\Phi((1+\sigma)\lambda_0)]^2} - 1. \quad (13)$$

Therefore, a regression model can be derived from the LSN distribution. In this model, θ is the mean of the response variable and $\psi_2(\sigma, \lambda)$ can be interpreted as a precision parameter in a way such that, for fixed θ, the larger the value of $\psi_2(\sigma, \lambda)$, the larger the variance of X.

Let us now consider that the random variable X_i is related to a vector of k covariates $\mathbf{y}_i = (y_{1i}, \ldots, y_{ki})'$ associated with the ith observation, where $i = 1, 2\ldots, n$. This vector assigns a weight of observable features is related to θ_i through a log link function and $\beta = (\beta_1, \ldots, \beta_k)'$ is a vector of regressors with $\beta_j \in \mathbb{R}$ for $j = 1, \ldots, k$. The regression model takes the form

$$\begin{aligned} X_i &\sim \mathrm{LSN}(\theta_i, \sigma, \lambda), \\ \log\theta_i &= \mathbf{y}_i'\boldsymbol{\beta}. \end{aligned} \quad (14)$$

Observe that log link ensures that θ_i falls within the interval $(0, \infty)$.

On the contrary, sometimes it is desirable to deal with a lower value in the range of X by shifting this random variable, say $X - \alpha$ with $\alpha > 0$, and therefore a shifted version of the previous distribution is obtained. In this case, it is convenient to work with the pdf given by

$$f_{\mu,\sigma,\alpha}(x) = \frac{\phi(\tilde{r}_{\mu,\sigma,\alpha}(x))}{\sigma(x-\alpha)}\frac{\Phi((1+\tilde{r}_{\mu,\sigma,\alpha}(x))\lambda)}{\Phi(\lambda_0)}, \quad x > \alpha, \quad (15)$$

where

$$\tilde{r}_{\mu,\sigma,\alpha}(x) = \frac{\log(x-\alpha) - \mu}{\sigma}. \quad (16)$$

Figure 1 shows pdf (15) as compared to the Pareto and PAT (Pareto ArcTan distribution) (this generalization of the
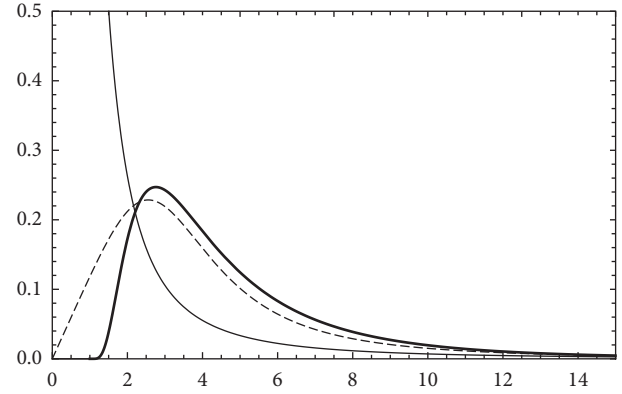

FIGURE 1: Graph of the Pareto (thin line), PAT (dashed line), and LSN (thick line) with the same mean given by (22).

Pareto distribution was proposed by Gómez-Déniz and Calderín-Ojeda [13]), all of them with the same mean, say (22), and a threshold of $\alpha = 1$. It is discernible that the tail of the LSN density seems larger than those ones of the other models.

By taking

$$\mu = -\frac{\sigma^2}{2} + \log\left[\frac{(\theta-\alpha)\Phi(\lambda_0)}{\Phi((1+\sigma)\lambda_0)}\right], \quad (17)$$

it is guaranteed that (15) has mean equal to $\theta > \alpha$. In this case, for the associated regression model, we will use the link function:

$$\theta_i = \alpha + \exp\{\mathbf{y}_i'\boldsymbol{\beta}\}, \quad (18)$$

which guarantees that θ_i falls within the interval $[\alpha, \infty)$.

It is already known (see Lemma 2 in [25]) that

$$\lim_{z \to \infty} \frac{-\ln\Phi(-z)}{z^2} = \frac{1}{2}. \quad (19)$$

Then, it is straightforward to see that if $\lambda < 0$, we have

$$\lim_{x \to \infty} \frac{-\ln\Phi(1 + r_{\mu,\sigma,\alpha}(x))}{[\lambda(1 + r_{\mu,\sigma,\alpha}(x))]^2} = \frac{1}{2}. \quad (20)$$

This can be used to easily obtain asymptotic values of the pdf given in (15). In fact, if $\lambda < 0$, we have the following approximation:

$$f(x) \approx \frac{\phi(r_{\mu,\sigma,\alpha}(x))}{\sigma(x-\alpha)\Phi(\lambda_0)}\exp\left[-\frac{1}{2}((1 + r_{\mu,\sigma,\alpha}(x))\lambda)^2\right], \quad x > \alpha, \quad x \to \infty. \quad (21)$$

3. Empirical Illustrations

In this section, the versatility of the proposed skew lognormal model, as compared with the Pareto distribution and the Pareto ArcTan Distribution [13], is tested using three datasets. The first dataset is the danishuni that can be found in the *R* package CASdatasets collected at *Copenhagen Reinsurance* and comprises 2167 fire losses over the period 1980 to 1990, adjusted for inflation to reflect 1985 values and

are expressed in millions of Danish Krone. The second dataset is norfire comprises 9181 fire losses over the period 1972 to 1992 from an unknown Norwegian company. A priority of 500 thousands of Norwegian Krone (NKR) (if this amount is exceeded, the reinsurer becomes liable to pay) was applied to obtain this dataset. The third dataset considers hospital costs in the state of Wisconsin. It can be found in the personal web page of Professor E. Frees. It examines the impact of several predictors on hospital charges obtained from the Office of Health Care Information, Wisconsin's Department of Health and Human Services, in the year 1989. The response variable is the *logarithm of total hospital charges per number of discharges*, and the explanatory variables considered are the *health service area, hospital discharge costs, the type of payer, size of the area population, number of hospital beds*, and *average income within the area*. For comparison reasons some alternative distributions have been considered in this paper, their pdf's are provided in Appendix A and B. Some descriptive statistics of the three variables of interest are shown in Table 1.

Parameter estimation for all the models considered in this paper has been carried out by the method of maximum likelihood using Mathematica v.12.0® and also confirmed by using WinRATS v.7.0. Both moments and the maximum likelihood methods are suitable to estimate the vector of parameters of the distribution via sample observations, as shown in Appendix A and B. Codes are available from the authors upon request. For details about software, see Ruskeepaa [26] and Brooks [27].

Since closed expressions are not available for the maximum likelihood estimates and the computation of the global maximum of the logarithm function of the likelihood is not guaranteed, thus it is advisable to use several seed points as the starting value. It is also prudent to use different optimization methods (Newton–Raphson, Broyden–Fletcher–Goldfarb–Sanno, BGGS) (this is an iterative method for solving unconstrained nonlinear optimization problems which can be seen in Broyden [28, 29] and Shanno [30]) that ensure that the same solution is obtained from any of these methods. The standard errors of the estimates can be calculated by inverting the Hessian matrix. In this sense, both Mathematica and WinRats have at least two methods to reach it. The first is to retrieve them from the Cholesky factors (this package is available on the web upon request). The second, faster, is to obtain them by finite differentiation. Furthermore, WinRats package also offers the possibility to directly compute the maximum of the log-likelihood providing the elements of the Fisher information matrix. In fact, for the examples considered in this section, these two packages were used to quickly compute the maximum likelihood estimates. Commands for fitting the skew normal and log-skew normal distributions are also available in stata (see [31]).

Parameter estimates and their standard errors (in brackets), negative of the maximum of the log likelihood function, and AIC for the three datasets considered are shown in Tables 2 and 3. Model assessment is derived through the following information criteria. Negative log likelihood (NLL) is calculated by taking the negative of the value of the log-likelihood evaluated at the ML estimates;

TABLE 1: Descriptive statistics of the dependent variable.

Data	Mean	Stand. dev.	Minimum	Maximum
Danish	56.339	534.161	1.00	14239
Norwegian	2217.21	7759.97	500.000	465365
Hospital cost (in log scale)	7.94537	0.75167	6.10504	9.64179

Akaike information criterion (AIC) is computed as twice the NLL and evaluated at the ML estimates, plus twice the number of estimated parameters; Consistent Akaike Information Criteria (CAIC), a corrected version of the AIC is proposed by Bozdogan [32] to overcome the tendency of the AIC overestimating the complexity of the underlying model as it lacks the asymptotic property of consistency. In order to calculate the CAIC, a correction factor based on the sample size is used to compensate for the overestimating nature of AIC. The CAIC is defined as twice the NLL plus $k(1 + \ln(n))$, where k is the number of free parameters and n refers to the sample size. Note that a model with a lower statistics value is preferred to one with a higher value. All these results are shown in Table 2. Furthermore, we also include the Kolmogorov–Smirnov test (KS) and the Anderson–Darling test (AD) to express the fit of the model to the data in terms of the distribution function. For these statistics, smaller values indicate a better fit of the model to the data. Note that they not only provide a way to measure the fit in terms of distribution functions but also allow us to perform hypothesis testing for model selection purposes. The p value of the test statistics, computed using the Monte Carlo method by using 1000 simulations, is shown in brackets. An extremely small p value may lead to a confident rejection of the null hypothesis that the data comes from the proposed model. It can be seen that LSN distribution is not rejected at the usual significance levels for both tests.

Graphs of histogram of the data and superimposed fitted densities are given in Figure 2. As can be seen, the proposed distribution provides a good fit for the empirical data.

In Table 4, the LSN distribution is fitted to the Wisconsin hospital costs' dataset. From left to right the parameter estimates, standard errors (SE), and the corresponding p values calculated based on the t-Wald statistic are displayed. Besides, AIC and BIC values for each model are provided in the last two rows of the table. The estimates of the regressors associated with the explanatory variables, *number of beds* and *income*, are not significant at the usual nominal level.

Finally, in order to choose a model that provides an acceptable description of the loss process for the Danish and Norwegian datasets, we should verify that the population limited expected values, computed numerically by

$$L(x) = E[\min(X, x)] = \int_0^x y \, dF(y) + x\overline{F}(x), \qquad (22)$$

are close to the empirical ones. As is well known, (22) is the expected amount per claim retained by the insured on a policy with a fixed amount deductible of x. In this case, the empirical limited expected value function was calculated based on $E_n(x) = (1/n)\sum_{i=1}^n \min(x_i, x)$. Obviously, when x

TABLE 2: Parameter estimates and their standard errors (in brackets), negative of the maximum of the log likelihood function, and AIC without covariates.

$\widehat{\lambda}$	$\widehat{\theta}$	$\widehat{\sigma}$	$\widehat{\alpha}$	NLL	AIC	CAIC	KS	AD
Danish data								
−1.324	3.121	2.075	0.993	3361.486	6730.970	6757.700	0.029	0.001
(0.308)	(0.097)	(0.222)	(0.004)				(0.319)	(0.136)
Norwegian data								
657.218	1894.500	1.239	318.836	20966.835	41941.700	41969.100	0.031	0.001
(18.987)	(43.876)	(0.027)	(9.600)				(0.148)	(0.049)

TABLE 3: Parameter estimates and standard errors (in brackets) of the homogeneous model together with some measures of model selection for different distributions considered for Wisconsin hospital costs' dataset.

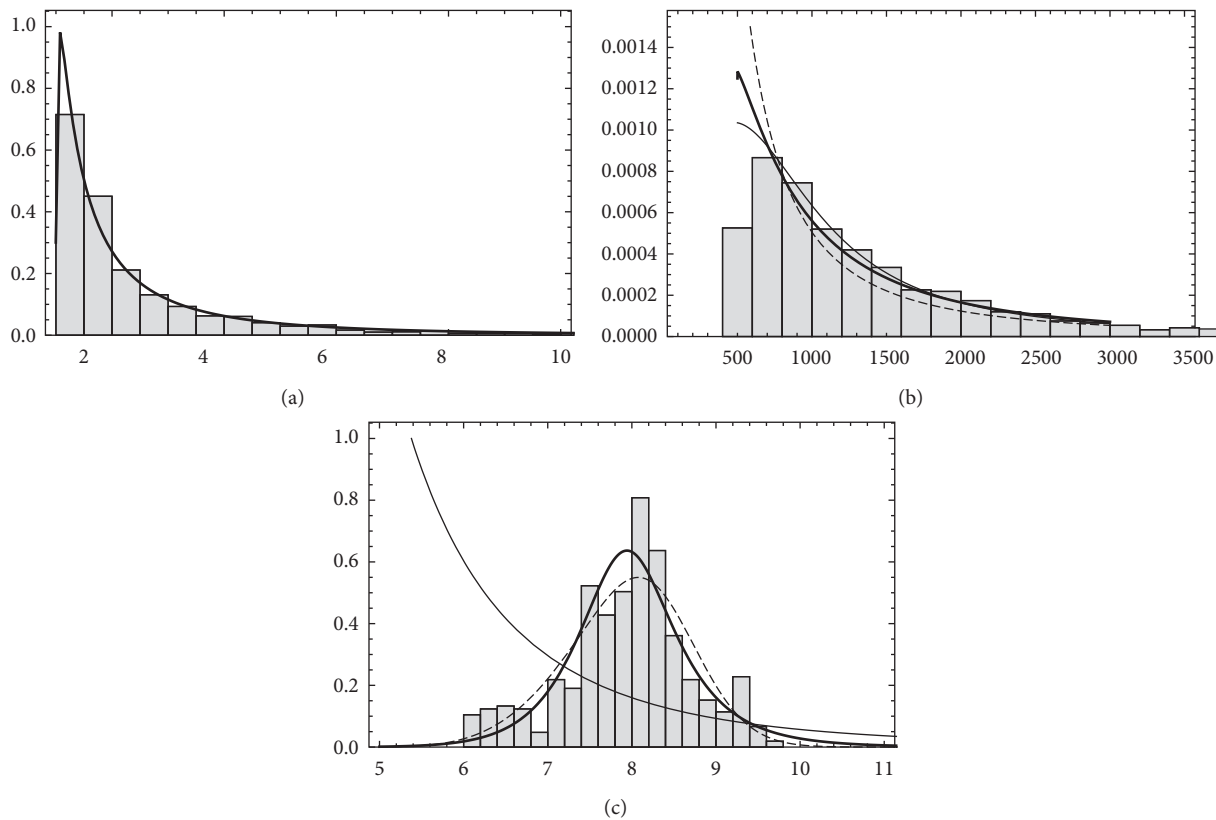| Model	$\widehat{\lambda}$	$\widehat{\theta}$	$\widehat{\sigma}$	$\widehat{\alpha}$	NLL	AIC	CAIC	KS	AD
Pareto			3.620	6.000	937.019	1876.040	1881.300	0.345	0.224
			(0.158)					(0.000)	(0.000)
PAT		15.900	647.273	5.307	600.407	1206.330	1222.120	0.846	1.355
		(0.628)	(10.897)	(0.083)				(0.000)	(0.000)
LSN	−3.132	7.941	0.277	2.601	590.134	1188.270	1209.330	0.051	0.004
	(0.338)	(0.031)	(0.007)	(0.123)				(0.492)	(0.260)

(a)

(b)

(c)

FIGURE 2: Smooth histogram of the empirical Danish, Norwegian insurance data, and hospital cost (in log scale) as compared to fitted models. Pareto (dashed line), PAT (thin line), and LSN (thick line). (a) Danish insurance loss data. (b) Norwegian insurance loss data. (c) Hospital costs' data.

tends to infinity, $L(x)$ and $E_n(x)$ approach $E(X)$ and the sample mean, respectively.

Table 5 below exhibits the limited expected value for different values of the policy limit x considered for the hospital costs' dataset. It is observed that the values obtained

from the LSN distribution adheres closely to the observed empirical limited expected values obtained from the Pareto and PAT distributions. Similarly, in Figure 3, empirical and fitted limited expected value function for this dataset and also for the Danish and Norwegian data are shown. As can

TABLE 4: Parameter estimates for Wisconsin hospital costs dataset including covariates.

Variable	Estimate	SE	t-Stat.	p value
Hospital discharge costs	0.015	0.002	6.37	0.000
Health service area	0.002	0.001	1.74	0.080
Payer	0.012	0.005	2.43	0.014
Size area population	0.015	0.005	3.10	0.001
Number of beds	−0.010	0.008	−1.20	0.228
Income	0.001	0.005	0.31	0.756
Constant	1.607	0.008	200.32	0.000
α	0.563	0.269	2.09	0.036
λ	−2.974	0.361	−8.21	0.000
σ	0.186	0.014	13.03	0.000

NLL = 564.619
AIC = 1149.240
CAIC = 1201.890

TABLE 5: :Limited expected value for the different distributions considered and different values of the fixed amount deductible x for hospital costs dataset.

	Limited expected value			
Deductible	Empirical	Pareto	PAT	LSN
6.00	6.00000	6.00000	5.99853	5.99810
6.15	6.14978	6.14346	6.14740	6.14661
6.30	6.29677	6.27480	6.29568	6.29419
6.45	6.44047	6.39529	6.44308	6.44044
6.60	6.58092	6.50605	6.58927	6.58482
6.75	6.71905	6.60806	6.73373	6.72666
6.90	6.85458	6.70218	6.87580	6.86517
7.05	6.98883	6.78918	7.01456	6.99940
7.20	7.11933	6.86973	7.14881	7.12827
7.35	7.24475	6.94442	7.27702	7.25059
7.50	7.36489	7.01380	7.39730	7.36506
7.65	7.47366	7.07833	7.50753	7.47039
7.80	7.57179	7.13843	7.60564	7.56533
7.95	7.65844	7.19450	7.69007	7.64882
8.10	7.73372	7.24686	7.76028	7.72019
8.25	7.79120	7.29582	7.81694	7.77925
8.40	7.83213	7.34167	7.86162	7.82640
8.55	7.86265	7.38465	7.89634	7.86260
8.70	7.88567	7.42498	7.92311	7.88928
8.85	7.90437	7.46287	7.94369	7.90813
9.00	7.91949	7.49850	7.95951	7.92088
9.15	7.93157	7.53205	7.97170	7.92914
9.30	7.94065	7.56367	7.98111	7.93427
9.45	7.94438	7.59349	7.98839	7.93731
9.60	7.94527	7.62165	7.99406	7.93905
9.75	7.94537	7.64826	7.99848	7.94001
9.90	7.94537	7.67342	8.00194	7.94051
10.00	7.94537	7.68945	8.00383	7.94070

be observed, the LSN distribution stays closer to the empirical limited expected values for the three datasets considered.

3.1. Out-of-Sample Validation of LSN Distribution.
We are interested now in showing the power of the LSN distribution to predict the number of claims out-of-sample via quantile-quantile plots (QQ-plots) of randomized quantile residuals.

For this reason, the dataset `danishuni` is randomly partitioned into two disjoint subsets of different sizes. The first subset (dataset A) is used for fitting the models, whereas the second subset (dataset B) is used for graphing the quantile-quantile plots of the randomized quantile residuals to check for normality. The residuals for the out-of-sample datasets are exhibited in Figure 4 for different in-sample sizes, i.e., size of dataset A. For comparison purposes, we have we have plotted the residuals for the whole dataset (top-left graph)
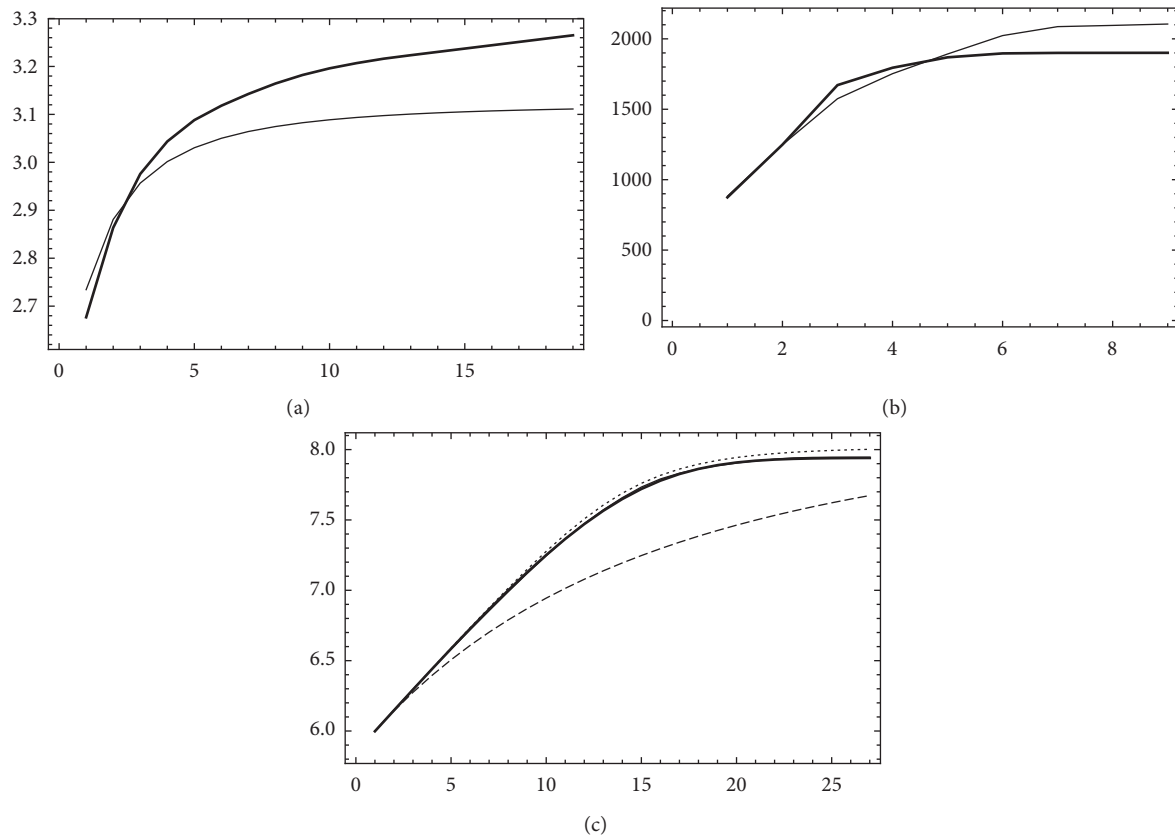
(a)

(b)

(c)

FIGURE 3: Empirical (thick line) and fitted limited expected values: Pareto distribution (dashed line), PAT distribution (dotted line), and LSN (thin line) for the (a) Danish data, (b) Norwegian data, and (c) hospital cost data.
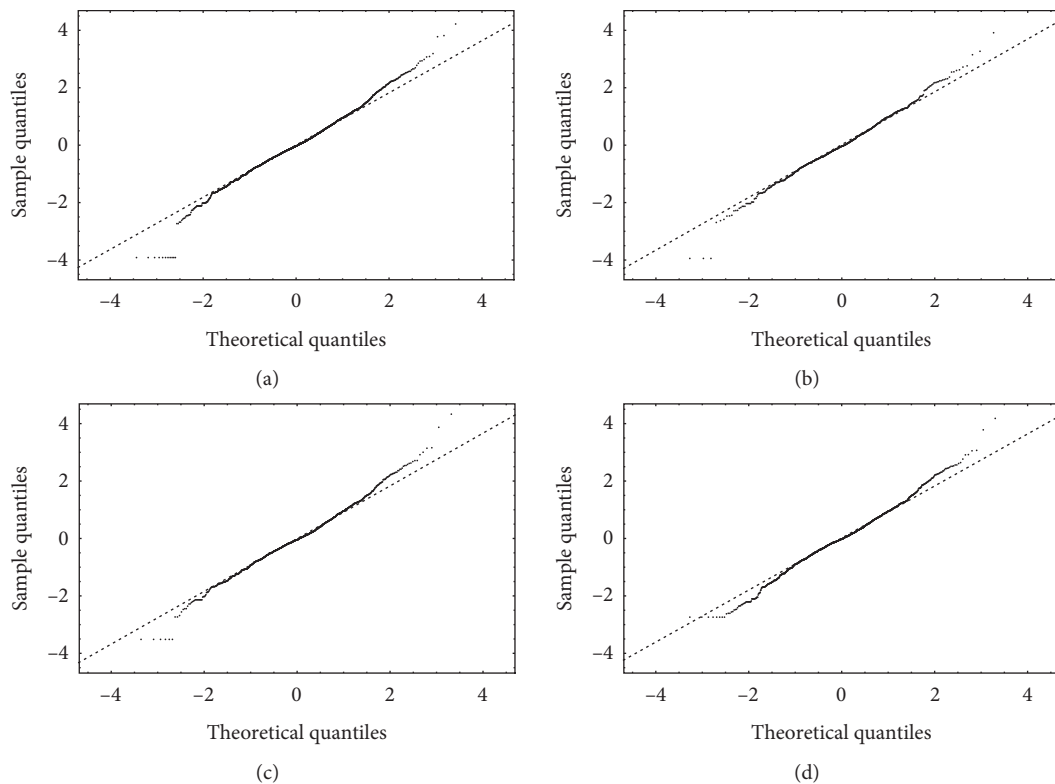

(a)

(b)

(c)

(d)

FIGURE 4: QQ-plots of randomized quantile residuals for the LSN distribution for the whole dataset `danishuni` and different in-sample sizes. (a) Whole dataset size-2167. (b) In-sample size-1500. (c) In-sample size-1000. (d) In-sample size-750.

with sample size 2167. A perfect alignment with the 45° line implies the residuals are normally distributed. It is observable that the residuals for LSN distribution adheres reasonably well to the line throughout the residual distribution. However, for all different in-sample sizes considered, the model tends to overestimate the lower quantiles and underestimate the upper quantiles. In general, the predictive power of the model is acceptable.

4. Conclusions and Extensions

In this paper, the use of the skew lognormal distribution has been proposed as a suitable model for describing claims' size empirical data. To the best of our knowledge, this probabilistic family has not been previously considered in the actuarial literature. The advantages of this distribution are twofold. On the one hand, it includes heavy tails which makes it an interesting model to describe severity data and; on the other hand, the distribution can be rewritten to allow for the incorporation of predictors to explain a response variable.

Finally, it is worthy to point out that the distribution introduced in this paper could be extended to other areas of risk theory. For example, the LSN distribution could be employed as the secondary distribution (the distribution of the claims size) in the compound collective risk model. In particular, this distribution could be used as an approximation of the distribution of the total claims' size instead of the traditionally considered normal distribution. Furthermore, calculation of reinsurance premiums based on this distribution is an issue that deserves to be studied.

Appendix

A. PDF of the Alternative Distributions

The pdf of the classical Pareto and PAT distributions used in this work are given by

$$f(x) = \frac{\sigma \alpha^{\sigma}}{x^{\sigma+1}}, \quad x \geq \alpha, \ \sigma > 0,$$

$$f(x) = \frac{1}{\tan^{-1} \sigma} \frac{\sigma \theta x^{\theta-1} \alpha^{\theta}}{\sigma \alpha^{\theta} + x^{2\theta}}, \quad x \geq \alpha, \ \theta > 0, \ \sigma > 0, \quad \text{(A.1)}$$

respectively.

B. Normal Equations

Let us firstly consider the case of the model without covariates and let $\tilde{x} = \{x_1, x_2, \ldots, x_n\}$ be a sample obtained from distribution (3). By denoting $\Theta = (\mu, \sigma, \lambda)$ as the vector of parameters to be estimated, the log-likelihood is proportional to

$$\ell(\tilde{x}; \Theta) = \sum_{i=1}^{n} \left[\log \Phi\big((1 + r_{\mu,\sigma}(x_i))\lambda\big) - \frac{1}{2} r_{\mu,\sigma}(x_i)^2 \right]$$
$$- n[\log \sigma - \log \Phi(\lambda_0)]. \quad \text{(B.1)}$$

The required normal equations to compute the maximum likelihood estimates are

$$\frac{\partial \ell(\tilde{x}; \Theta)}{\partial \mu} = \sum_{i=1}^{n} r_{\mu,\sigma}(x_i) - \lambda \sum_{i=1}^{n} \frac{\phi\big((1 + r_{\mu,\sigma}(x_i))\lambda\big)}{\Phi\big((1 + r_{\mu,\sigma}(x_i))\lambda\big)} = 0,$$

$$\frac{\partial \ell(\tilde{x}; \Theta)}{\partial \sigma} = -\lambda \sum_{i=1}^{n} \frac{r_{\mu,\sigma}(x_i)\phi\big((1 + r_{\mu,\sigma}(x_i))\lambda\big)}{\Phi\big((1 + r_{\mu,\sigma}(x_i))\lambda\big)} + \sum_{i=1}^{n} r_{\mu,\sigma}(x_i)^2 - n = 0, \quad \text{(B.2)}$$

$$\frac{\partial \ell(\tilde{x}; \Theta)}{\partial \lambda} = \frac{n\lambda_0 \phi(\lambda_0)}{\Phi(\lambda_0)} - \lambda(1 + \lambda^2) \sum_{i=1}^{n} \frac{(1 + r_{\mu,\sigma}(x_i))\phi\big((1 + r_{\mu,\sigma}(x_i))\lambda\big)}{\Phi\big((1 + r_{\mu,\sigma}(x_i))\lambda\big)} = 0.$$

Data Availability

The csv data used to support the findings of this study can be found in the R repository CASdatasets and also at https://instruction.bus.wisc.edu/jfrees/jfreesbooks/Regression%20Modeling/BookWebDec2010/data.html.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] A. McNeil, "Estimating the tails of loss severity distribution using extreme value theory," *ASTIN Bulletin*, vol. 27, no. 1, pp. 117–137, 1988.

[2] J. Beirlant and J. L. Teugels, "Modeling large claims in non-life insurance," *Insurance: Mathematics and Economics*, vol. 11, no. 1, pp. 17–29, 1992.

[3] J. Beirlant, J. Teugels, and P. Vynckier, *Practical Analysis of Extreme Values*, Leuven University Press, Leuven, Belgium, 1996.

[4] P. Embrechts, S. I. Resnick, and G. Samorodnitsky, "Extreme value theory as a risk management tool," *North American Actuarial Journal*, vol. 3, no. 2, pp. 30–41, 1999.

[5] H. Albrecher, J. Beirlant, and J. Teugels, *Reinsurance: Actuarial and Statistical Aspects*, Wiley, Hoboken, NJ, USA, 2017.

[6] B. Arnold, *Pareto Distributions*, International Cooperative Publishing House, Silver Spring, MD, USA, 1983.

[7] V. Brazauskas and R. Serfling, "Favorable estimators for fitting Pareto models: a study using goodness-of-fit measures with actual data," *ASTIN Bulletin*, vol. 33, no. 2, pp. 365–381, 2003.

[8] M. Rytgaard, "Estimation in the Pareto distribution," *ASTIN Bulletin*, vol. 20, no. 2, pp. 201–216, 1990.

[9] A. V. Boyd, "Fitting the truncated Pareto distribution to loss distributions," *Journal of the Staple Inn Actuarial Society*, vol. 31, pp. 151–158, 1988.

[10] A. J. Mata, "Pricing excess of loss reinsurance with reinstatements," *ASTIN Bulletin*, vol. 30, no. 2, pp. 349–368, 2000.

[11] J. M. Sarabia and F. Prieto, "The Pareto-positive stable distribution: a new descriptive model for city size data," *Physica A: Statistical Mechanics and Its Applications*, vol. 388, no. 19, pp. 4179–4191, 2009.

[12] E. Gómez-Déniz and E. Calderín-Ojeda, "A suitable alternative to the Pareto distribution," *Hacettepe Journal of Mathematics and Statistics*, vol. 43, no. 5, pp. 843–860, 2014.

[13] E. Gómez-Déniz and E. Calderín-Ojeda, "Modelling insurance data with the Pareto arctan distribution," *ASTIN Bulletin*, vol. 45, no. 3, pp. 639–660, 2015.

[14] M. Ghitany, E. Gómez-Déniz, and S. Nadarajah, "A new generalization of the Pareto distribution and its application to insurance data," *Journal of Risk and Financial Management (Special Issue: Extreme Values and Financial Risk)*, vol. 11, no. 10, pp. 1–14, 2008.

[15] N. Henze, "A probabilistic representation of the skew-normal distribution," *Scandinavian Journal of Statistics*, vol. 13, no. 4, pp. 271–275, 1986.

[16] B. Arnold and R. Beaver, "Skewed multivariate models related to hidden truncation and/or selective reporting (with discussion)," *Test*, vol. 11, no. 1, pp. 1–54, 2002.

[17] A. Azzalini and A. Valle, "The multivariate skew-normal distribution," *Biometrika*, vol. 83, no. 4, pp. 715–726, 1996.

[18] A. Azzalini and A. Capitanio, "Statistical applications of the multivariate skew normal distribution," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 61, no. 3, pp. 579–602, 1999.

[19] A. Azzalini, *The Skew-Normal and Related Families*, Cambridge University Press, Cambridge, UK, 2013.

[20] G. D. Lin and J. Stoyanov, "The logarithmic skew-normal distributions are moment-indeterminate," *Journal of Applied Probability*, vol. 46, no. 3, pp. 909–916, 2009.

[21] A. Azzalini, D. Cappello, and S. Kotz, "Log-skew-normal and log-skew-t distributions as model for family income data," *Journal of Income Distribution*, vol. 11, pp. 12–20, 2003.

[22] H. Bolfarine, H. W. Gómez, and L. I. Rivas, "The log-bimodal-skew-normal model. A geochemical application," *Journal of Chemometrics*, vol. 25, no. 6, pp. 329–332, 2011.

[23] O. Venegas, H. Bolfarine, D. I. Gallardo, A. Vergara-Fernández, and H. W. Gómez, "A note on the log-alpha-skew-normal model with geochemical applications," *Applied Mathematics & Information Sciences*, vol. 10, no. 5, pp. 1697–1703, 2016.

[24] D. B. Owen, "Tables for computing bivariate normal probabilities," *The Annals of Mathematical Statistics*, vol. 27, no. 4, pp. 1075–1090, 1956.

[25] B. Arnold and G. Lin, "Characterizations of the skew-normal and generalized chi distributions," *Sankhya*, vol. 66, pp. 593–606, 2004.

[26] H. Ruskeepaa, *Mathematica Navigator. Mathematics, Statistics, and Graphics*, Academic Press, Cambridge, MA, USA, Third edition, 2009.

[27] C. Brooks, *RATS Handbook to Accompany Introductory Econometrics for Finance*, Cambridge University Press, Cambridge, UK, 2009.

[28] C. G. Broyden, "A class of methods for solving nonlinear simultaneous equations," *Mathematics of Computation*, vol. 19, no. 92, pp. 577–593, 1965.

[29] C. G. Broyden, "Quasi-Newton methods and their application to function minimisation," *Mathematics of Computation*, vol. 21, no. 99, pp. 368–381, 1967.

[30] D. F. Shanno, "Conditioning of quasi-Newton methods for function minimization," *Mathematics of Computation*, vol. 24, no. 111, pp. 647–657, 1970.

[31] Y. V. Marchenko and M. G. Genton, "A suite of commands for fitting the skew-normal and skew-t models," *The Stata Journal: Promoting Communications on Statistics and Stata*, vol. 10, no. 4, pp. 507–539, 2010.

[32] H. Bozdogan, "Model selection and Akaike's Information Criterion (AIC): the general theory and its analytical extensions," *Psychometrika*, vol. 52, no. 3, pp. 345–370, 1987.