Computational and Applied Mathematics (2022) 41:171
https://doi.org/10.1007/s40314-022-01875-6

A novel claim size distribution based on a Birnbaum–Saunders and gamma mixture capturing extreme values in insurance: estimation, regression, and applications

Emilio Gómez–Déniz[1] · Víctor Leiva[2] · Enrique Calderín–Ojeda[3] · Christophe Chesneau[4]

Received: 3 November 2021 / Revised: 5 March 2022 / Accepted: 6 April 2022

Abstract
Data including significant losses are a pervasive issue in general insurance. The computation of premiums and reinsurance premiums, using deductibles, in situations of heavy right tail for the empirical distribution, is crucial. In this paper, we propose a mixture model obtained by compounding the Birnbaum–Saunders and gamma distributions to describe actuarial data related to financial losses. Closed-form credibility and limited expected value premiums are obtained. Moment estimators are utilized as starting values in the non-linear search procedure to derive the maximum-likelihood estimators and the asymptotic variance–covariance matrix for these estimators is determined. In comparison to other competing models commonly employed in the actuarial literature, the new mixture distribution provides a satisfactory fit to empirical data across the entire range of their distribution. The right tail of the empirical distribution is essential in the modeling and computation of reinsurance premiums. In addition, in this paper, to make advantage of all available data, we create a regression structure based on the compound distribution. Then, the response variable is explained as a function of a set of covariates using this structure.

Keywords Actuarial data · Discrete mixture distribution · Mathematica software · Moment and maximum-likelihood estimation

Mathematics Subject Classification 62E15 · 62F10 · 62P05 · 91B05 · 91G70

Communicated by Eduardo Souza de Cursi.

✉ Christophe Chesneau
christophe.chesneau@gmail.com

1 Department of Quantitative Methods in Economics and TiDES Institute, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain

2 School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

3 Department of Economics, University of Melbourne, Melbourne, Australia

4 Department of Mathematics, Université de Caen Normandie, Caen, France

Published online: 11 May 2022

1 Introduction

In insurance and reinsurance, large claims modeling is a hot topic, especially in mathematical risk theory. In this regard, the conventional Pareto distribution has largely been considered as a good claim size distribution (see Arnold 1983). Furthermore, there are numerous instances where the empirical distribution of data exhibits light or heavy asymmetry. This is often the case with actuarial and financial data, which frequently contain long (heavy) tails indicating the presence of extreme values (see Beirlant et al. 1996; Embrechts et al. 1999, among others). For a comprehensive study about reinsurance, see the book by Albrecher et al. (2017).

A suitable claim size distribution in credibility and reinsurance premium calculations plays an important role in the rating process. The use of the extreme-value methodology in reinsurance has been examined by Beirlant et al. (2005). Traditionally, the models considered in the extreme-value methodology are non-parametric or semi-parametric in nature. One of the limitations of the extreme value approach is that it only focuses on the last part (right tail) of the distribution. However, the actuary could also be interested in modeling the central part of the distribution to calculate reinsurance premiums, since it flexibly manages the different layers. From a practical perspective, it is often convenient to utilize a fully parametric model to assess the portfolio risk in the tails and central part of the claim size distribution. Therefore, heavy-tailed parametric models provide a fair trade-off between the statistical flexibility and computational simplicity to calculate insurance and reinsurance premiums.

Based on this motivation, the objective of the present study is to formulate a new family of probability distributions with a domain in the positive real numbers. This new family is obtained by mixing the Birnbaum–Saunders (BS) with gamma distributions. From the pioneering work presented by Birnbaum and Saunders (1969) in the context of fatigue life problems, many papers have examined the BS distribution in the applied statistical literature based on different scenarios ranging from business to industry. For an extensive review of this model, the reader is led to the book by Leiva (2016). Recently, in analyzing loss data, Hashemi et al. (2019a, b) and Naderi et al. (2020a, b) introduced some BS-based distributions for analyzing positive support financial data. The excellent results obtained encourage further development and extension of the BS distribution in this applied direction, which is one aim of the present study.

The rest of this paper is organized as follows. In Sect. 2, the proposed distribution is exhaustively analyzed, and several of its properties are stated. Some appealing actuarial results in the context of risk theory are shown in Sect. 3. Numerical applications are provided in Sect. 4. Concluding remarks and ideas for future research are discussed in Sect. 5.

2 The new mixture model

After combining the BS and gamma distributions, this section introduces the probability density function (PDF) of the new mixture distribution. For both the PDF and related moments, closed-form formulations are provided.

2.1 The Birnbaum–Saunders distribution

To begin, we recall that a continuous random variable X follows a BS distribution with parameters of shape $\theta > 0$ and scale $\sigma > 0$, that is, $X \sim \mathrm{BS}(\theta, \sigma)$, if its PDF is given by

$$g(x;\theta,\sigma)=\frac{\sqrt{\theta}(1+x\sigma)}{2\sqrt{x^3\sigma}}\phi\left(\frac{\sqrt{\theta}(x\sigma-1)}{\sqrt{x\sigma}}\right),\quad x>0, \tag{1}$$

where ϕ represents the standard normal PDF, formulated as $\phi(z)=(1/\sqrt{2\pi})\exp(-z^2/2)$, for $z\in\mathbb{R}$. The mean and variance of X are, respectively, stated as

$$\mathrm{E}(X)=\frac{1+2\theta}{2\sigma\theta},\quad \mathrm{Var}(X)=\frac{5+4\theta}{4(\sigma\theta)^2}, \tag{2}$$

while the cumulative distribution function (CDF) is expressed by $G(x)=\bar{\Phi}((1-x\sigma)\sqrt{\theta/(x\sigma)})$, for $x>0$, where $\bar{\Phi}(z)=1-\Phi(z)$ is the standard normal survival function and Φ is its CDF.

Definition 1 We say that the random variable X follows a BS-gamma distribution if it admits the following stochastic representation: $(X|\Theta=\theta)\sim\mathrm{BS}(\theta,\sigma)$, with $\Theta\sim\mathcal{G}(\alpha,\beta)$, for $\theta>0$, where $\mathcal{G}(\alpha,\beta)$ is the gamma distribution with parameters of shape $\alpha>0$ and rate $\beta>0$. From now on, a random variable X following this distribution is denoted by $X\sim\mathrm{BSG}(\sigma,\alpha,\beta)$.

2.2 The Birnbaum–Saunders gamma distribution

The PDF of the BSG distribution, that is obtained from the stochastic representation stated in Definition 1, can be obtained by compounding it as

$$\begin{aligned} f(x;\sigma,\alpha,\beta)&=\frac{1+x\sigma}{2\sqrt{2\pi}\sqrt{x^3\sigma}}\frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty\theta^{\alpha+1/2-1}\exp\left(-\theta\left(\beta+\frac{(x\sigma-1)^2}{2x\sigma}\right)\right)\mathrm{d}\theta\\ &=\frac{(1+x\sigma)\beta^\alpha(2x\sigma)^{\alpha+1/2}}{B(\alpha,1/2)(2x)^{3/2}\sigma^{1/2}\left(2x\sigma\beta+(x\sigma-1)^2\right)^{\alpha+1/2}}\\ &=\frac{(2\beta x\sigma)^\alpha(1+x\sigma)}{2xB(\alpha,1/2)(1+(x\sigma)^2-2\bar{\beta}x\sigma)^{\alpha+1/2}},\quad x>0, \end{aligned} \tag{3}$$

with Γ denoting the complete gamma function, $\Gamma(1/2)=\sqrt{\pi}$, B representing the complete beta function, and $\bar{\beta}=1-\beta$.

Figure 1 shows the PDF expressed in (3) for different values of its parameters. Note that, in all cases, the BSG distribution has different features related to positive skewness, heavy tails, and unimodality. The first and third features are found in actuarial data related to claim size or damage amount. The second one is crucial to model extreme values and, therefore, to compute reinsurance premiums.

The unconditional mean and second-order moment of $X\sim\mathrm{BSG}(\sigma,\alpha,\beta)$ are, respectively, given by

$$\mathrm{E}(X)=\mathrm{E}\left[\mathrm{E}(X\mid\Theta)\right]=\frac{1}{\sigma}+\frac{\beta}{2\sigma(\alpha-1)},\quad \alpha>1, \tag{4}$$

$$\mathrm{E}(X^2)=\mathrm{E}\left[\mathrm{E}(X^2\mid\Theta)\right]=\frac{4+2\alpha(\alpha-3)+\beta(3\beta+4\alpha-8)}{2\sigma^2(\alpha-1)(\alpha-2)},\quad \alpha>2. \tag{5}$$

Now, combining (4) and (5), we get the variance of the BSG distribution stated as

$$\mathrm{Var}(X)=\frac{\beta\left(4(2-\beta)+\alpha(4(\alpha-3)+5\beta)\right)}{4(\alpha-2)(\alpha-1)^2\sigma^2},\quad \alpha>2. \tag{6}$$

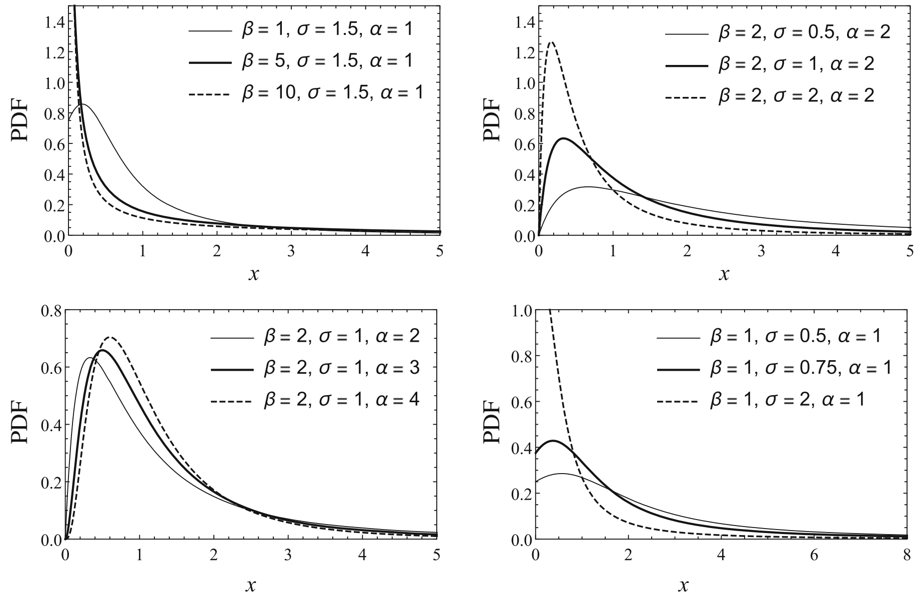

Fig. 1 Plots of the BSG PDF for the indicated values of its parameters

The upper kth moment of $X \sim \mathrm{BSG}(\sigma, \alpha, \beta)$ can be obtained by compounding raw moments of the BS PDF formulated in (1) and the gamma PDF as

$$\mathrm{E}(X^k) = \frac{\sqrt{\theta}\exp(\theta)}{2\sigma^k\sqrt{2\pi}}\left(K_{k-\frac{1}{2}}(\theta) + K_{k+\frac{1}{2}}(\theta)\right),$$

where

$$K_\nu(u) = \left(\frac{u}{2}\right)^\nu \frac{\sqrt{\pi}}{\Gamma(\nu+\frac{1}{2})}\int_1^\infty (\tau^2-1)^{\nu-1/2}\exp(-u\tau)\,\mathrm{d}\tau,$$

which denotes the Bessel function of the second kind of order ν and argument u.

Coefficients of skewness (asymmetry) and kurtosis of the BSG distribution are plotted in Fig. 2. Note that the values of the GBS coefficients of skewness nor kurtosis change with the value of the parameter σ. Therefore, they do not depend on the value of σ. To plot the corresponding graphs, we choose a fixed value of σ to calculate these coefficients as a function of α and β. Observe that both coefficients decrease with the values of α and β for a fixed value of σ.

A value x from $X \sim \mathrm{BSG}(\sigma, \alpha, \beta)$ can be generated via the following algorithm:

- Generate a random number u from the standard uniform distribution, U(0,1) in short.
- Obtain a value θ from a gamma distributed random variable Θ with shape parameter α and rate parameter β.
- Fix values for σ, α, and β of $X \sim \mathrm{BSG}(\sigma, \alpha, \beta)$.
- Calculate

$$x = \frac{1}{4\sigma^2}\left(-\sqrt{\frac{\sigma}{\theta}}\Phi^{-1}(1-u) + \sqrt{\frac{\sigma}{\theta}(\Phi^{-1}(1-u))^2 + 4\sigma}\right)^2,$$

where Φ^{-1} is the quantile function of the standard normal distribution or its inverse CDF.

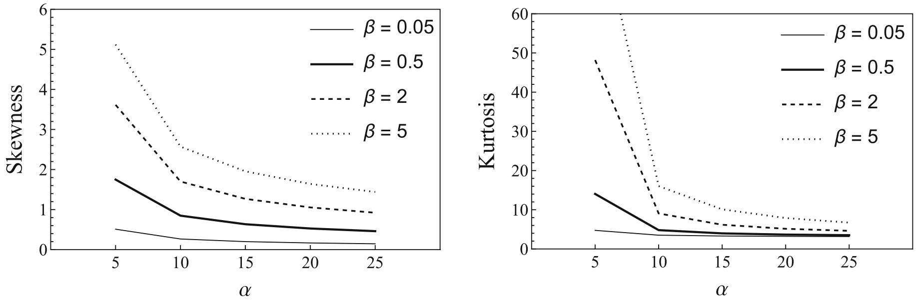

Fig. 2 Plots of coefficients of skewness (left panel) and kurtosis (right panel) of the BSG distribution for the indicated values of the parameters α and β for a fixed value of $\sigma = 2$

2.3 Experience rating via credibility

In an insurance portfolio, the Bayesian approach is advantageous in situations where the inherent risk is not considered a single entity, but as a combination of similar risks that share common characteristics. However, the different groups can be viewed as different regimes of the entire risk. For example, when observing the number of claims or the claim size in a fire insurance portfolio, the different types of services offered by the insurer are an element that must be certainly considered. It is notorious that the fire risk rate increases with the size of the insured object. In this case, the Bayesian methodology offers an elegant and helpful solution, mainly when the data are limited and the practitioner's prior knowledge is relevant.

It is worth mentioning that both the gamma and BS distributions are conjugate, that is, the posterior distribution follows the same distribution as the prior but with updated parameters, with the BS PDF being provided in (1). Then, it should be included in the conjugate distributions catalog.

Next, we focus on the case of the gamma distribution and derive its posterior distribution when this is employed as the prior distribution. Such a result is useful in experience rating when credibility premiums are required.

Proposition 1 *Let $X_1, \ldots, X_n$ be independent and identically distributed random variables following the PDF given in* (1)*. Let us suppose that Θ follows a prior gamma distribution with PDF denoted as $\pi(\theta)$ and parameters $\alpha > 0$, $\beta > 0$, that is, $\pi(\theta) \propto \theta^{\alpha-1} \exp(-\beta\theta)$. Then, the posterior distribution of Θ given the sample $(X_1, \ldots, X_n)$ is $\mathcal{G}(\alpha^*, \beta^*)$, where*

$$\alpha^* = \alpha + \frac{n}{2}, \tag{7}$$

$$\beta^* = \beta + \sum_{i=1}^{n} \frac{(X_i\sigma - 1)^2}{2X_i\sigma}. \tag{8}$$

Proof After applying the Bayes theorem and arranging parameters, we arrive at the final result. □

Proposition 1 is helpful for providing credibility premiums in the insurance context. Consider the random variable X, which represents the amount of an insurance contract claimed by policyholders and is BS distributed with the PDF given in (1). In actuarial statistics, the expression defined in (2), now represented as $P(\theta)$, is referred to as the risk premium obtained when the squared-error loss function is considered (see Gómez-Déniz 2008). The risk (unknown) parameter of a specific risk is called θ in this situation.

Let us suppose now that the parameter θ is considered random, varying between the policyholders, and reflecting the heterogeneity of the insurance portfolio. Suppose also that this parameter follows a prior distribution, also known as the structure function, with PDF $\pi(\theta) \propto \theta^{\alpha-1}\exp(-\beta\theta)$. Then, using Proposition 1 and taking again the squared-error loss function, the collective premium, $P = \mathrm{E}_{\pi(\theta)}[P(\theta)]$ namely, results as in the expression defined in (4) and the Bayes premium is obtained from this latter by updating parameters according to (7)–(8). Hence, we have that

$$P^* = Z(n)h(\tilde{x}) + (1 - Z(n))P, \tag{9}$$

where

$$h(\tilde{x}) = \frac{1}{n\sigma}\sum_{i=1}^{n}\left(X_i\sigma + \frac{1}{2X_i\sigma}\right) \tag{10}$$

and

$$Z(n) = \frac{n}{n + 2(\alpha - 1)}. \tag{11}$$

In actuarial statistics, the term stated in (11) is referred to as the credibility factor. In this case, the Bayesian premium given in (9) is written as a convex sum of the collective premium and a function of the data provided in (10). For more details on the credibility theory, see Bühlmann and Gisler (2005).

2.4 A simplified submodel

As an example, when $\alpha = 1$, we get a mixture of the BS and exponential distributions. This distribution lacks interest as it does not have a mean value. For this reason, it is more interesting to consider the special case assuming that $\beta = 1$, which gives a two-parameter distribution with PDF formulated as

$$f(x; \sigma, \alpha) = \frac{(2\sigma x)^{\alpha} g_1(x; \sigma)}{2x B(\alpha, 1/2)\,(g_2(x; \sigma))^{\alpha+1/2}}, \quad x > 0. \tag{12}$$

Henceforward, for a random variable X that follows the PDF expressed in (12), we write $X \sim \mathrm{BSGT}(\sigma, \alpha)$. Simple computations show that this distribution is unimodal, with the mode (or modal value) being the solution of the equation with respect to x stated as

$$\frac{1}{g_1(x; \sigma)} - \frac{1 + 2\alpha}{g_2(x; \sigma)} - (1 + \alpha) = 0.$$

2.5 Parameters' estimation

Let us consider a sample $\tilde{X} = (X_1, \ldots, X_n)^\top$ of size n obtained from the PDF defined in (12), with observed values $\tilde{x} = (x_1, \ldots, x_n)^\top$. Moment estimates for σ and α can be generated easily from the expressions formulated in (4) and (6). The moment estimate of the parameter α is the solution of the equation with respect to α given by

$$4 + \alpha(4(\alpha - 3) + 5) = \frac{s^2}{\bar{x}^2}(\alpha - 2)(2\alpha - 1)^2,$$

where $\bar{x}$ and s^2 are the sample mean and variance, respectively. Once the estimate of α is obtained, say $\widehat{\alpha}$, the estimate of σ is computed as $\widehat{\sigma} = (2\widehat{\alpha} - 1)/(2\bar{x}(\widehat{\alpha} - 1))$. These estimates can be used as starting values for the non-linear search procedure shown below.

To proceed with the maximum-likelihood method, the log-likelihood function, $\ell(\tilde{x}; \tilde{\varsigma})$ namely, based on the data $\tilde{x}$ collected from the PDF stated in (3), is proportional to

$$\begin{aligned}\alpha\left(n\left(\log(\beta) + \log(2\sigma)\right) + \sum_{i=1}^{n}\log(x_i)\right) + \sum_{i=1}^{n}\log(g_1(x_i;\sigma))\\ -n\log(B(\alpha, 1/2)) - \left(\alpha + \frac{1}{2}\right)\sum_{i=1}^{n}\log(g_2(x_i;\sigma) - 2\sigma\bar{\beta}x_i),\end{aligned} \tag{13}$$

where $\tilde{\varsigma} = (\sigma, \alpha, \beta)^\top$ denotes the vector of parameters to be estimated. From (13), we derive the estimation equations given by

$$\begin{aligned}\frac{\partial\ell(\tilde{x};\tilde{\varsigma})}{\partial\sigma} &= \frac{n\alpha}{\sigma} + \sum_{i=1}^{n}\frac{x_i}{g_1(x_i;\sigma)} - (2\alpha+1)\sum_{i=1}^{n}\frac{x_i(x_i\sigma - \bar{\beta})}{g_2(x_i;\sigma) - 2\sigma\bar{\beta}x_i} = 0,\\ \frac{\partial\ell(\tilde{x};\tilde{\varsigma})}{\partial\alpha} &= n\left(\psi\left(\alpha + \frac{1}{2}\right) - \psi(\alpha) + \log(2\sigma) + \log(\beta)\right)\\ &\quad + \sum_{i=1}^{n}\log(x_i) - \sum_{i=1}^{n}\log\left(g_2(x_i;\sigma) - 2\sigma\bar{\beta}x_i\right) = 0,\\ \frac{\partial\ell(\tilde{x};\tilde{\varsigma})}{\partial\beta} &= \frac{n\alpha}{\beta} - (2\alpha+1)\sigma\sum_{i=1}^{n}\frac{x_i}{g_2(x_i;\sigma) - 2\sigma\bar{\beta}x_i} = 0,\end{aligned}$$

where $\psi(s) = \mathrm{d}\log(\Gamma(s))/\mathrm{d}s$ is the digamma function. The corresponding second partial derivatives are established as

$$\begin{aligned}\frac{\partial\ell(\tilde{x};\tilde{\varsigma})}{\partial\alpha^2} &= n\left(\psi_1\left(\alpha + \frac{1}{2}\right) - \psi_1(\alpha)\right),\\ \frac{\partial\ell(\tilde{x};\tilde{\varsigma})}{\partial\alpha\partial\sigma} &= \frac{n}{\sigma} + 2\bar{\beta}\sum_{i=1}^{n}\frac{x_i}{g_2(x_i;\sigma) - 2\sigma\bar{\beta}x_i},\\ \frac{\partial\ell(\tilde{x};\tilde{\varsigma})}{\partial\alpha\partial\beta} &= \frac{n}{\beta} - 2\sigma\sum_{i=1}^{n}\frac{x_i}{g_2(x_i;\sigma) - 2\sigma\bar{\beta}x_i},\\ \frac{\partial\ell(\tilde{x};\tilde{\varsigma})}{\partial\sigma^2} &= -\frac{n\alpha}{\sigma^2} - \sum_{i=1}^{n}\left(\frac{x_i}{g_1(x_i;\sigma)}\right)^2 - (2\alpha+1)\sum_{i=1}^{n}\frac{x_i^2(g_2(x_i;\sigma) - 2\bar{\beta}^2)}{(g_2(x_i;\sigma) - 2\sigma\bar{\beta}x_i)^2},\\ \frac{\partial\ell(\tilde{x};\tilde{\varsigma})}{\partial\sigma\partial\beta} &= -(2\alpha+1)\sum_{i=1}^{n}\frac{x_i(g_2(x_i;\sigma) - 2x_i\sigma^2)}{(g_2(x_i;\sigma) - 2\sigma\bar{\beta}x_i)^2},\\ \frac{\partial\ell(\tilde{x};\tilde{\varsigma})}{\partial\beta^2} &= -\frac{n\alpha}{\beta^2} + 2(2\alpha+1)\sigma^2\sum_{i=1}^{n}\left(\frac{x_i}{g_2(x_i;\sigma) - 2\sigma\bar{\beta}x_i}\right)^2,\end{aligned}$$

where ψ_1 is the derivative of the digamma function.

For the special case $\beta = 1$, the corresponding Fisher information matrix has a closed-form expression given by

$$\mathcal{J}(\widehat{\sigma}, \widehat{\alpha}) = \begin{pmatrix} \mathcal{J}_{11} & \mathcal{J}_{12} \\ \mathcal{J}_{21} & \mathcal{J}_{22} \end{pmatrix},$$

where

$$\begin{aligned} \mathcal{J}_{11}(\widehat{\sigma}, \widehat{\alpha}) &= n\left(\psi_1(\widehat{\alpha}) - \psi\left(\widehat{\alpha} + \frac{1}{2}\right)\right), \\ \mathcal{J}_{12}(\widehat{\sigma}, \widehat{\alpha}) = \mathcal{J}_{21}(\widehat{\sigma}, \widehat{\alpha}) &= 0, \\ \mathcal{J}_{22}(\widehat{\sigma}, \widehat{\alpha}) &= \frac{n}{\widehat{\sigma}^2}\left(\widehat{\alpha} + \frac{2\widehat{\alpha}(1+\widehat{\alpha})}{3+2\widehat{\alpha}} - \frac{1+2\widehat{\alpha}}{2}\right. \\ &\quad \left. + \frac{1}{4}\left(\frac{\pi\sqrt{2}\csc(\pi\widehat{\alpha})}{B(\widehat{\alpha}, 1/2)} + 2\sum_{j=1}^{2}(-1)^{j-1}\mathcal{D}(j; \widehat{\sigma}, \widehat{\alpha}),\right)\right), \end{aligned}$$

with

$$\mathcal{D}(j; \widehat{\sigma}, \widehat{\alpha}) = \frac{\widehat{\alpha} + j - 1}{\widehat{\alpha} - j}\, {}_2F_1\left(1, \frac{\widehat{\alpha} + j + 1}{2}; \frac{2 + j - \widehat{\alpha}}{2}; -1\right),$$

and ${}_2F_1$ being the hypergeometric function defined as

$${}_2F_1(a, b; c; z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}\, \mathrm{d}t.$$

The asymptotic covariance matrix of $(\widehat{\sigma}, \widehat{\alpha})$ is obtained by inverting the information matrix.

2.6 Simulation experiment

Now, we perform some simulation experiments using the bootstrap method (see Efron and Tibshirani 1993; Wilcox 2010, among others), to examine the behavior of the maximum-likelihood estimators of the BSG parameters and the correlation between these estimators. In this study, we utilize the `Mathematica` software to generate values of a BSG distributed random variable employing the PDF stated in (3) based on the aforementioned algorithm.

The sample sizes used in this simulation study to compute the estimates are $n = 500$ and $n = 750$. The average estimates and square root of the mean squared errors (MSEs) are calculated based on 1000 replicates. Additional replicates seem to be unnecessary, as computational time would be prohibitive, even though less replicates might reduce the statistical accuracy obtained. Clearly, as the sample size increases, the bias and MSE tend to decrease, which seems to empirically verify the consistency properties of the maximum-likelihood estimators. These results, together with the correlation ρ between the parameter estimators for the BSG distribution, of the averaged simulated estimates are reported in Table 1.

Table 1 Empirical estimate, standard error, bias, MSE, and correlation between the parameter estimators for the BSG distribution based on the indicated values

$n=500$	$\sigma=0.5$	$\alpha=2$	$\beta=2$	$\sigma=2$	$\alpha=0.5$	$\beta=1$
Estimate	0.461	2.060	2.095	1.858	0.511	1.045
Standard error	0.022	0.455	0.649	0.185	0.041	0.212
Bias	−0.040	0.059	0.095	−0.141	0.011	0.045
MSE	0.002	0.210	0.430	0.054	0.002	0.047
Correlation	$\rho_{\widehat{\sigma},\widehat{\alpha}}=-0.093$	$\rho_{\widehat{\sigma},\widehat{\beta}}=-0.122$	$\rho_{\widehat{\alpha},\widehat{\beta}}=0.961$	$\rho_{\widehat{\sigma},\widehat{\alpha}}=-0.068$	$\rho_{\widehat{\sigma},\widehat{\beta}}=-0.092$	$\rho_{\widehat{\alpha},\widehat{\beta}}=0.795$
$n=750$	$\sigma=0.5$	$\alpha=2$	$\beta=2$	$\sigma=2$	$\alpha=0.5$	$\beta=1$
Estimate	0.513	1.890	1.818	1.953	0.486	0.951
Standard error	0.020	0.304	0.467	0.148	0.031	0.178
Bias	0.013	−0.109	−0.181	−0.047	−0.013	−0.049
MSE	5.6×10^{-4}	0.104	0.251	0.024	0.001	0.034
Correlation	$\rho_{\widehat{\sigma},\widehat{\alpha}}$ =-0.018	$\rho_{\widehat{\sigma},\widehat{\beta}}=0.024$	$\rho_{\widehat{\alpha},\widehat{\beta}}=0.958$	$\rho_{\widehat{\sigma},\widehat{\alpha}}=0.174$	$\rho_{\widehat{\sigma},\widehat{\beta}}=0.274$	$\rho_{\widehat{\alpha},\widehat{\beta}}=0.798$
$n=500$	$\sigma=1$	$\alpha=3$	$\beta=2$	$\sigma=3$	$\alpha=2$	$\beta=5$
Estimate	0.992	3.318	2.285	3.010	1.808	4.527
Standard error	0.041	1.420	1.279	0.224	0.320	1.205
bias	−0.007	0.319	0.285	0.010	−0.191	−0.472
MSE	0.002	2.125	1.716	0.050	0.139	1.675
Correlation	$\rho_{\widehat{\sigma},\widehat{\alpha}}=0.045$	$\rho_{\widehat{\sigma},\widehat{\beta}}=0.037$	$\rho_{\widehat{\alpha},\widehat{\beta}}=0.989$	$\rho_{\widehat{\sigma},\widehat{\alpha}}=-0.175$	$\rho_{\widehat{\sigma},\widehat{\beta}}=-0.194$	$\rho_{\widehat{\alpha},\widehat{\beta}}=0.942$
$n=750$	$\sigma=1$	$\alpha=3$	$\beta=2$	$\sigma=3$	$\alpha=2$	$\beta=5$
Estimate	0.997	2.868	1.820	3.047	1.934	5.065
Standard error	0.031	0.735	0.607	0.181	0.298	1.165
Bias	−0.002	−0.132	−0.179	0.046	−0.066	0.065
MSE	9.9^{-4}	0.558	0.400	0.035	0.093	1.360
Correlation	$\rho_{\widehat{\sigma},\widehat{\alpha}}=0.058$	$\rho_{\widehat{\sigma},\widehat{\beta}}=0.087$	$\rho_{\widehat{\alpha},\widehat{\beta}}=0.980$	$\rho_{\widehat{\sigma},\widehat{\alpha}}=0.003$	$\rho_{\widehat{\sigma},\widehat{\beta}}=0.007$	$\rho_{\widehat{\alpha},\widehat{\beta}}=0.952$

3 Further results

In this section, the CDF, equilibrium distribution, hazard rate, and excess-of-loss reinsurance for the BSG distribution are introduced.

3.1 Cumulative distribution function

Some additional essential results in risk theory are also obtained in closed-form expressions for the PDF of the two-parameter distribution stated in (12). The CDF of this distribution can be written in terms of the hypergeometric function as

$$F(x;\sigma,\alpha)=\frac{(2\sigma x)^{\alpha}}{2B(\alpha,1/2)}\sum_{j=1}^{2}\frac{(\sigma x)^{j-1}}{\alpha+j-1}\mathcal{H}_j(x;\sigma,\alpha), \tag{14}$$

where

$$\mathcal{H}_j(x;\sigma,\alpha)={}_2F_1\left(\frac{j(\alpha+1)-1}{2},\frac{\alpha}{j}+\frac{1}{2};\frac{\alpha+j+1}{2};-(\sigma x)^2\right),\quad j\in\{1,2\}.$$

Therefore, we have that

$$F(x;\sigma,\alpha)=\int_0^x\frac{(2t\sigma)^{\alpha}}{2B(\alpha,1/2)t}\frac{g_1(t;\sigma)}{(g_2(t;\sigma))^{\alpha+1/2}}\,\mathrm{d}t.$$

Now, using the changing variable $t=\tau x$, we get

$$F(x;\sigma,\alpha)=\int_0^1\frac{(2\tau x\sigma)^{\alpha}}{2B(\alpha,1/2)\tau}\frac{1+\sigma x\tau}{(1+(\sigma x\tau)^2)^{\alpha+1/2}}\,\mathrm{d}\tau,$$

which is directly related to the hypergeometric function and

$$\int z^m\,{}_2F_1(a,b,c,-(\sigma z)^2)\,\mathrm{d}z=\frac{z^{m+1}}{(m+1)}{}_3F_2\left(\left(a,b,\frac{m+1}{2}\right),\left(c,\frac{m+3}{2}\right),-(\sigma z)^2\right). \tag{15}$$

Note that the expression defined in (15) can be verified by differentiating the right-hand side of the equation to achieve the formula under the integral. Then, the result follows after some algebra.

3.2 Equilibrium distribution

The equilibrium distribution, also known as integrated tail distribution, is a common probability model in insurance (see Rolski et al. 1999; Konstantinides 2018). The integrated tail distribution of a random variable X with CDF F is given by

$$F_I(x)=\frac{1}{\mathrm{E}(X)}\int_0^x\bar{F}(y)\,\mathrm{d}y.$$

The integrated tail distribution can be calculated in closed-form for the BSG distribution, as shown in the following result.

Proposition 2 *The integrated tail distribution of the CDF given in* (14) *is provided by*

$$F_I(x) = \frac{2\sigma(\alpha-1)}{2\alpha-1}\Big(x - \frac{2^{\alpha-1}}{\sigma}\sum_{j=1}^{2}\Big(\frac{\sigma x}{2}\Big)^{\alpha+j}\mathcal{N}_1(j;\sigma,\alpha)\mathcal{N}_2(j;\sigma,\alpha)\Big), \tag{16}$$

where $\alpha > 1/2$*, with* $r_j^k(\alpha) = (\alpha + j + k)/2$*, for* $k \in \{1, 2\}$*, and*

$$\begin{aligned}\mathcal{N}_1(j;\sigma,\alpha) &= \frac{\Gamma(\alpha+1/2)\Gamma(2r_j^{-1}(\alpha))}{\Gamma(\alpha)\Gamma(r_j^1(\alpha))\Gamma(r_j^2(\alpha))},\\ \mathcal{N}_2(j;\sigma,\alpha) &= {}_3F_2\Big(\Big(\frac{\alpha}{j}+\frac{1}{2}, \frac{j(\alpha+1)-1}{2}, r_j^0(\alpha)\Big), \Big(r_j^1(\alpha), r_j^2(\alpha)\Big), -(\sigma x)^2\Big).\end{aligned}$$

Proof The necessary result can be obtained using the expressions defined in (2), (14), and (15), as well as some algebra. □

Let us now consider the risk model (see Rolski et al. 1999; Konstantinides 2018, among others) and assume a positive security loading, λ namely, for the claim size distribution with regularly varying tails. Then, by utilizing the formula stated in (16), we can get a rough approximation of the probability of ruin provided by

$$\Psi(u) \sim \frac{1}{\lambda}\bar{F}_I(u), \quad u \to \infty,$$

where $\bar{F}_I(u) = 1 - F_I(u)$ and u is the initial surplus level.

3.3 Mean residual life

The failure rate of the integrated tail distribution is defined by $\gamma_I(x) = \bar{F}(x)/\int_x^\infty \bar{F}(y)\,\mathrm{d}y$. It may also be calculated in closed form for the BSG distribution. Furthermore, the reciprocal of $\gamma_I(x)$ is the mean residual life, which can be expressed as

$$e(x) = \mathrm{E}(X - x | X > x) \tag{17}$$

$$= x\left(\frac{1+\alpha}{x\sigma(1+\alpha)\mathcal{M}_1(\alpha) + \alpha\mathcal{M}_2(\alpha)}\left(\mathcal{M}_1(\alpha) - \frac{\alpha\sigma x\mathcal{M}_3(\alpha)}{\bar{\alpha}}\right) - 1\right), \tag{18}$$

where

$$\begin{aligned}\mathcal{M}_1(\alpha) &= {}_2F_1\left(\frac{\alpha}{2}, \alpha+\frac{1}{2}; \frac{\alpha}{2}+1; -\frac{1}{(\sigma x)^2}\right),\\ \mathcal{M}_2(\alpha) &= {}_2F_1\left(\alpha+\frac{1}{2}, \frac{\alpha+1}{2}; \frac{\alpha+3}{2}; -\frac{1}{(\sigma x)^2}\right),\\ \mathcal{M}_3(\alpha) &= {}_2F_1\left(-\frac{\bar{\alpha}}{2}, \alpha+\frac{1}{2}; \frac{\alpha+1}{2}; -\frac{1}{(\sigma x)^2}\right),\end{aligned}$$

with $\bar{\alpha} = 1 - \alpha$. Note that the mean residual life defined in (18) is not a linear function of x, as is the case with the Pareto distribution.

In general insurance, it is critical to utilize a heavy right-tailed distribution. The Pareto and log-normal distributions have been used to predict losses in motor third-party liability insurance, fire insurance, and catastrophe insurance, among other applications. We recall that any probability distribution, specified through its CDF $F(x)$ on the real line, is heavy right-tailed if and only if, for every $t > 0$, $\mathrm{e}^{tx}\bar{F}(x)$ diverges as x tends to $+\infty$, where $\bar{F}(x) =$

$1-F(x)$. Thus, the distribution decays to zero slower than the exponential distribution, since for any fixed t > 0 (see Rolski et al. 1999), it is verified that $\bar{F}(x+t) \sim \bar{F}(x)$, as $x \to \infty$. Since a long-tailed distribution is likewise heavy right-tailed, the BSG distribution presented in this paper is heavy right-tailed as well.

The regular variation is another major topic in extreme-value theory (see Rolski et al. 1999; Konstantinides 2018, among others). The following definition formalizes this concept.

Definition 2 A CDF is called regular varying at infinity with index $-\varrho$ if

$$\lim_{x\to\infty} \frac{\bar{F}(\tau x)}{\bar{F}(x)} = \tau^{-\varrho}, \quad \tau > 0,$$

where the parameter $\varrho \geq 0$ is called the tail index.

The theorem that follows establishes that the CDF defined in (14) is a regular variant of the Lebesgue measure.

Theorem 1 *The CDF associated with the PDF given in* (3) *is regular varying at infinity with index* $-\alpha$.

Proof It is obvious that, at the first glance, we get the indetermine form stated as

$$\lim_{x\to\infty} \frac{\bar{F}(\tau x;\sigma,\alpha,\beta)}{\bar{F}(x;\sigma,\alpha,\beta)} = \frac{0}{0}, \quad \tau > 0.$$

Now, by applying the L'Hospital rule, we get directly that

$$\lim_{x\to\infty} \frac{\bar{F}(\tau x;\sigma,\alpha,\beta)}{\bar{F}(x;\sigma,\alpha,\beta)} = \lim_{x\to\infty} \frac{\tau f(\tau x;\sigma,\alpha,\beta)}{f(x;\sigma,\alpha,\beta)} = \tau^{-\alpha},$$

and hence, the proof of the theorem is complete. □

An immediate consequence of Theorem 1 is the following corollary (see Jessen and Mikosch 2006).

Corollary 1 *Let* $X_1, \ldots, X_n$ *be independent and identically distributed as the random variable* X*, with common survival function given by* $\bar{F}(x) = 1 - F(x)$ *and* $S_n = \sum_{i=1}^{n} X_i$*, for* $n \geq 1$*. Then, we have that*

$$\Pr(S_n > x) \sim \Pr(X > x), \quad as \quad x \to \infty.$$

As a consequence of Corollary 1, we have that, if $P_n = \max_{i\in\{1,\ldots,n\}} X_i$, for $n \geq 1$, then $\Pr(S_n > x) \sim n \Pr(X > x) \sim \Pr(P_n > x)$. This means that, for enough large x, the event $\{S_n > x\}$ is due to the event $\{P_n > x\}$. As a result, exceedance of high thresholds by the sum S_n are caused by this threshold being exceeded by the largest value in the sample.

3.4 Excess-of-loss reinsurance

Among all the local types of reinsurance contracts, excess-of-loss reinsurance is the best one from the insurance perspective, since it minimizes the variability of the retained claims. As pointed out by Boland (2007), excess-of-loss reinsurance with a common retention level, say M, does not affect the frequency or rate of claims for the ceding company or insurer. However, it reduces the amount paid on larger claims. The best actuarial tool for analyzing excess-of-loss reinsurance is the limited expected value (LEV) function. For a non-negative claim, X namely, the LEV function is the expected value of the random variable $Y = \min\{X, M\}$. This quantity is presented in the following proposition for the BSG distribution.

Proposition 3 *Let X be a random variable denoting the size of an individual claim, with values only for claims greater than M. Assume that X has the PDF defined in (3). Then, the LEV function given $M > 0$ is stated as*

$$L(M) = \mathcal{T}(M) + M\bar{F}(M; \sigma, \alpha, \beta), \tag{19}$$

where

$$\mathcal{T}(M) = \frac{2^{\alpha-1} M (M\sigma)^{\alpha}}{B(\alpha, 1/2)} \sum_{j=1}^{2} \frac{(M\sigma)^{j-1}}{\alpha + j} \mathcal{P}(j; \sigma, \alpha),$$

for $\alpha > 1$, with

$$\mathcal{P}(j; \sigma, \alpha) = {}_2F_1\left(\frac{j^2 + 2\alpha}{2}, \frac{\alpha j + 1}{2}; \frac{\alpha + 2 + j}{2}; -(M\sigma)^2\right).$$

Proof The LEV function (see Boland 2007, p. 113) is defined as

$$L(M) = \mathrm{E}\left[\min(X, M)\right] = \int_0^M x f(x; \sigma, \alpha, \beta) \mathrm{d}x + M\bar{F}(M; \sigma, \alpha, \beta).$$

It is well known that

$$\frac{\mathrm{d}}{\mathrm{d}y} {}_2F_1(a, b; z; y) = \frac{ab}{z} {}_2F_1(1 + a, 1 + b; 1 + z; y),$$

from which note that $\int y f(y; \sigma, \alpha, \beta)\, \mathrm{d}y = \mathcal{T}(y)/\mathrm{d}y$. Now, we have that

$$L(M) = \mathcal{T}(M) - \mathcal{T}(0) + M\bar{F}(M; \sigma, \alpha, \beta),$$

but $\mathcal{T}(0) = 0$. Hence, we get the result. □

We have summarized the methodology presented in this work in the flowchart illustrated in Fig. 3.

4 Empirical illustrations

In this section, the proposed model's adaptability is demonstrated in comparison to other commonly used statistical models. We employ three well-known data sets from the actuarial literature for this aim.

4.1 The data sets

The first data set is the `danishuni`, which was collected at Copenhagen Reinsurance and can be obtained in the `R` package `CASdatasets`. It includes 2167 fire losses from 1980 to 1990, which have been adjusted for inflation to represent 1985 values and expressed in millions of Danish Krone. The actuarial literature has extensively examined this data set (see Albrecher et al. 2017).

The second data set comes from the US Agency of Health Research and Quality's Medical Expenditure Panel Survey (MEPS). The MEPS is a probability survey that offers nationally representative values of US civilian health-care usage, expenditures, payment sources, and insurance coverage. This data set can be found on Professor E. Frees' website (Wisconsin

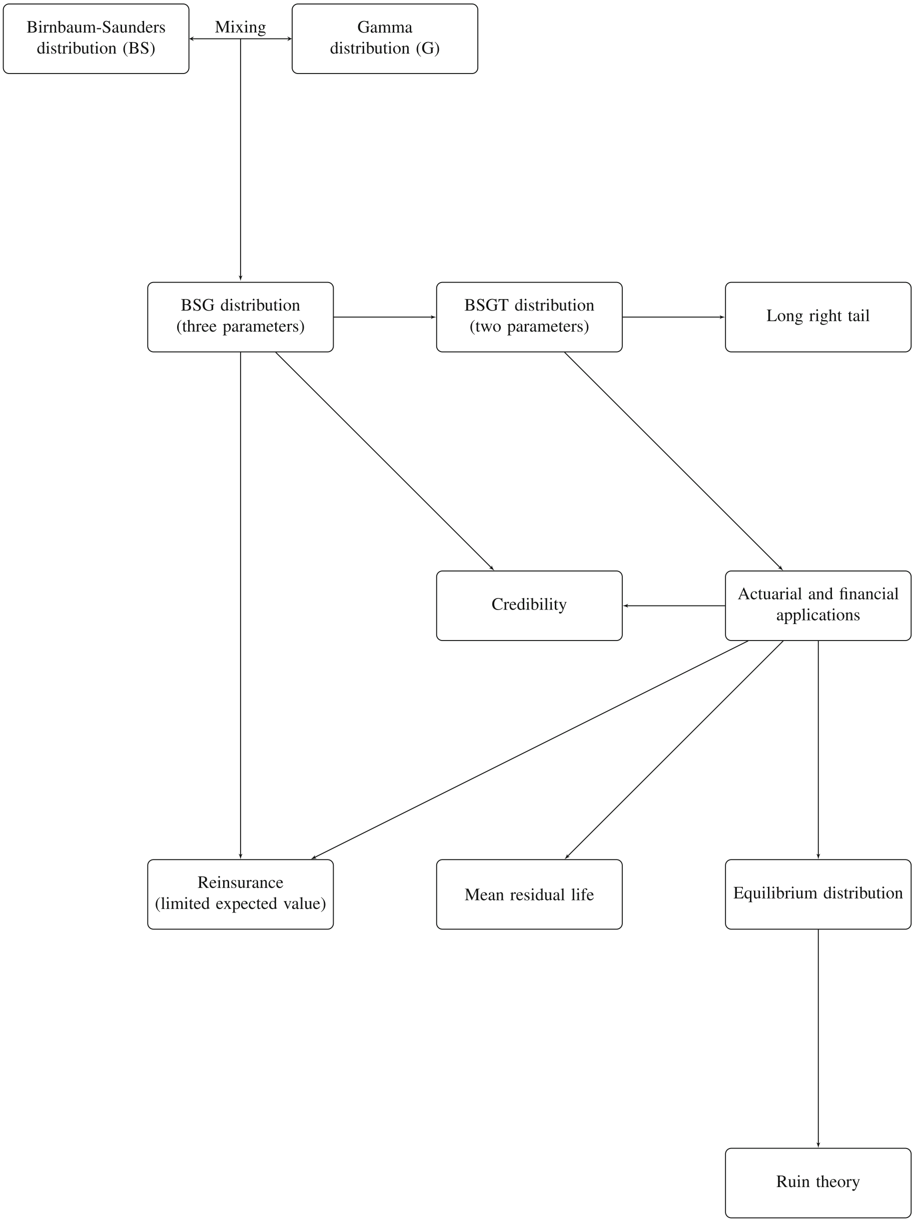

Fig. 3 Flowchart showing the methodology proposed

School of Business Research). This survey gathers thorough data on individuals from each medical care episode, including physician office visits, hospital emergency department visits, hospital outpatient visits, hospital inpatient stays, all other medical provider visits, and prescription drug use. These precise data enable the development of health-care models that can be employed to forecast future spending. We take the logarithm of the data from 157 people who had positive inpatient spending visits from this group.

Table 2 Descriptive statistics of the dependent variable for the indicated data set

Data set	Mean	Standard deviation	Minimum	Maximum
Danish	56.339	534.161	1.00	14239
MEPS	8.5653	1.29338	3.2700	13.3176
ABI	5.9535	33.1362	0.0050	1067.6970

The Insurance Research Council (IRC), a part of the American Institute for Chartered Property Casualty Underwriters and the Insurance Institute of America, provides the third data set, which is based on automobile bodily injury (ABI) claims. This data set is available in the `R` package `CASdatasets`, as well as on Professor E. Frees' website. The data set, which was gathered in 2002, has 1340 records that include demographic data regarding claimants, attorney involvement, and economic losses (in thousands of US dollars), among other variables.

Some descriptive statistics of the three response variables of interest for each data set are shown in Table 2.

4.2 Data analysis

We compare the BSG and BSGT distributions with other competitive distributions in the literature. As a benchmark, we consider the following distributions with the indicated PDF:

- Generalized gamma (GG) distribution (the gamma distribution is obtained for $\beta = 1$):

$$f(x;\sigma,\alpha,\beta) = \frac{\beta}{\alpha^{\sigma}\Gamma(\sigma/\beta)} x^{\sigma-1} \exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right], \quad x > 0,\ \sigma > 0,\ \alpha > 0,\ \beta > 0.$$

- Lognormal (LN) distribution:

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

- Inverse Gaussian (IG) distribution:

$$f(x;\sigma,\alpha) = \sqrt{\frac{\sigma}{2\pi x^3}} \exp\left[-\frac{\sigma}{2x\alpha^2}(x-\alpha)^2\right], \quad x > 0,\ \sigma > 0,\ \alpha > 0.$$

The parameters for the first two data sets were estimated with the maximum-likelihood method using `Mathematica`® `v.12.0` and also confirmed by `WinRATS v.7.0`.

As often considered in the actuarial literature, we employ a threshold $\delta = 1$, that is, we make the change of variable $Y = X + \delta$, for the Danish data (see Table 3). Both moment and maximum-likelihood methods are suitable for estimating the vector of the distribution parameters via sample observations. The authors can provide codes and data upon request. For details about the software, see Ruskeepaa (2009) and Brooks (2009). Parameter estimates, standard errors, the negative value of the maximized log-likelihood function (NLL), and values of the goodness-of-fit Kolmogorov–Smirnov (KS) and Anderson–Darling (AD) test statistics are reported in Tables 3 and 4 for the first two data sets. Smaller values indicate a better fit for the model to the data. For the KS and AD tests, the p-values computed using the Monte Carlo method with 2000 replicates being reported in parentheses. The null hypothesis that the data come from the specified model may be rejected if the p-value is small enough

Table 3 Parameters' estimates and standard errors (in parentheses), NLL, KS, and AD statistics and p-values (in parentheses) for the Danish data set

Distribution	$\widehat{\sigma}$	$\widehat{\alpha}$	$\widehat{\beta}$	δ	NLL	KS	AD
BSG	1.2685	0.8887	1.0329	1.00	3337.497	0.0175	0.00061
	(0.0379)	(0.0444)	(0.1016)	–	–	(0.8929)	(0.4080)
BSGT	1.2679	0.8761	–	1.00	3337.554	0.0170	0.000639
	(0.038)	(0.0219)	–	–	–	(0.9102)	(0.4225)
Gamma	0.5409	4.4113	–	1.00	3697.561	0.1306	0.0275
	(0.0136)	(0.1703)	–	–	–	(< 0.001)	(< 0.001)
IG	0.1056	2.3861	–	1.00	4565.750	0.3535	0.18131
	(0.0030)	(0.2435)	–	–	–	(< 0.001)	(< 0.001)
GG	2.2073	4.6E−5	0.1974	1.00	3377.197	0.1306	0.0275
	(0.0931)	(4.3E−6)	(0.0077)	–	–	(< 0.001)	(< 0.001)

Table 4 Parameters' estimates and standard errors (in parentheses), NLL, KS, and AD statistics and p-values (in parentheses) for the MEPS data set

Distribution	$\widehat{\sigma}$	$\widehat{\alpha}$	$\widehat{\beta}$	NLL	KS	AD
BSG	0.1159	1.4061	0.0135	252.412	0.051	0.0048
	(0.0011)	(0.3379)	(0.0052)	–	(0.9869)	(0.8710)
BSGT	0.1179	34.898	–	280.635	0.1337	0.0280
	(0.0016)	(3.9100)	–	–	(0.1205)	(0.0720)
Gamma	37.0214	0.2313	-	275.039	0.1274	0.02364
	(4.0904)	(0.0256)	–	–	(0.1564)	(0.1170)
LN	0.1746	2.1341	–	283.904	0.1465	0.03269
	(0.0093)	(0.0139)	–	–	(0.0688)	(0.0555)
IG	269.738	8.5653	–	285.964	0.15287	0.03598
	(30.4531)	(0.1222)	–	–	(0.0510)	(0.0230)
GG	9.8464	7.9415	5.2931	261.308	0.095	0.01011
	(1.6157)	(0.7171)	(0.9471)	–	(0.4707)	(0.4870)

(usually less than 5%). Observe that the proposed distributions are not rejected at the standard significance levels for both tests. Histograms of these two sets of data and superimposed fitted PDFs are displayed in Figure 4. Note that the proposed distributions provide an excellent fit to the empirical data in both cases.

Now, we must select a model that adequately describes the loss process for the Danish data set. We check that the theoretical limited anticipated values, obtained numerically using the formula given in (19), roughly match the empirical ones. The formula stated in (19) indicates the expected amount per claim retained by the insured on a policy with a set amount deductible of x, as it is well known. In this scenario, $\mathrm{E}_n(x) = (1/n)\sum_{i=1}^{n}\min(x_i, x)$ is used to calculate the empirical limited expected value function.

Table 7 reports the limited expected value for various x values examined with the Danish data set. Note that the values obtained from the two proposed distributions adhere closely to the empirical limited expected values rather than the benchmark distributions.

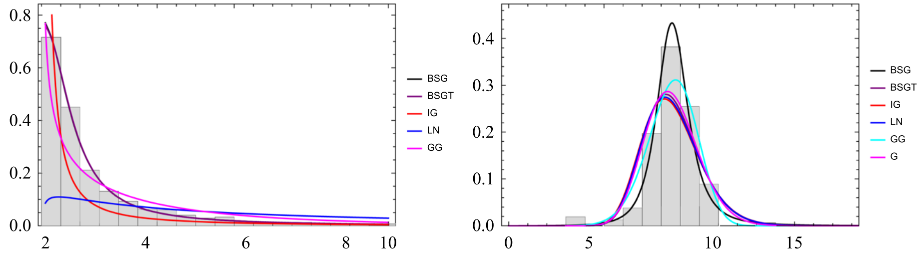

Fig. 4 Histogram of the Danish (left) and MEPS (right) data sets and superimposed PDFs of the indicated models

Table 5 Empirical and theoretical limited expected value for the indicate distribution and different values of the policy limit x with the Danish data set

Policy limit (x)	Distribution					
	Empirical	Gamma	GG	IG	BSGT	BSG
1.5	1.40345	1.39014	1.39237	1.26385	1.40522	1.40486
3.0	1.98083	2.13098	2.00091	1.57541	1.97954	1.97988
4.5	2.25875	2.56758	2.31136	1.75819	2.25069	2.25213
6.0	2.42371	2.84286	2.50320	1.89373	2.42953	2.43149
7.5	2.53644	3.02158	2.63318	2.00273	2.56424	2.56635
9.0	2.62514	3.13960	2.72643	2.09424	2.67294	2.67498
10.5	2.70067	3.21844	2.79603	2.17318	2.76442	2.76625
12.0	2.76495	3.27154	2.84955	2.24255	2.84361	2.84515
13.5	2.81840	3.30754	2.89166	2.30437	2.91358	2.91477
15.0	2.86364	3.33206	2.92541	2.36004	2.97634	2.97715
16.5	2.90339	3.34882	2.95289	2.41060	3.03333	3.03373
18.0	2.93800	3.36033	2.97556	2.45683	3.08557	3.08553
19.5	2.96744	3.36825	2.99446	2.49935	3.13384	3.13335
21.0	2.99151	3.37371	3.01038	2.53864	3.17872	3.17778
22.5	3.01245	3.37748	3.02390	2.57510	3.22069	3.21930
24.0	3.03150	3.38010	3.03548	2.60905	3.26013	3.25827
25.5	3.04898	3.38191	3.04545	2.64078	3.29734	3.29502
27.0	3.06406	3.38317	3.05410	2.67050	3.33257	3.32979
28.5	3.07718	3.38404	3.06163	2.69842	3.36604	3.36280
30.0	3.08811	3.38465	3.06824	2.72470	3.39793	3.39423

4.3 Regression analysis

In this subsection, we derive regression models associated with the BSG and BSGT distributions. Then, we apply these models to the MEPS data set when some of the covariates available in the data description file on the web page of Professor E. Frees are included.

To derive the regression models, let the parameter σ, defined in (3), be written as a function of the mean μ, that is, as

$$\sigma = \frac{1}{\mu}\left(1 + \frac{\beta}{2(\alpha - 1)}\right).$$

Table 6 Parameter estimates and p-values for the MEPS data set including covariates

Variable	BSGT		BSG	
	Estimate	p-value	Estimate	p-value
Age	0.018	0.709	0.100	< 0.01
Anylimit	0.008	0.813	− 0.023	0.289
College	− 0.065	0.296	0.466	0.215
Highsch	−0.037	0.359	0.219	0.244
Gender	0.005	0.868	− 0.026	0.231
Mnhpoor	− 0.021	0.619	0.000	0.969
Insure	0.122	0.030	0.093	< 0.01
Usc	0.004	0.905	− 0.003	0.849
Unemploy	0.029	0.511	0.018	0.345
Managedcare	0.021	0.581	0.028	0.120
Famsize	− 0.004	0.650	− 0.004	0.362
Countip	0.052	< 0.01	0.039	< 0.01
Countop	− 3.136	< 0.01	− 0.328	< 0.01
Race	0.007	0.588	0.008	0.287
Region	0.006	0.674	0.007	0.350
Education	0.001	0.953	− 0.245	0.191
Status	0.017	0.327	0.002	0.761
Income	0.017	0.194	0.000	0.972
Hstatus	0.007	0.611	0.001	0.829
α	45.393	< 0.01	1.060	< 0.01
β			0.005	< 0.01
Constant	− 8.037	< 0.01	1.585	< 0.01
NLL	260.141		222.066	
Observations	157		157	

By doing this, covariates can be implemented into the model to explain the response variable Y as a function of their observed values $\tilde{x} = (x_{i1}, \dots, x_{ik})^\top$, with $i \in \{1, \dots, n\}$. Then, we write $\mu \equiv \mu(\tilde{x}, \tilde{\eta}) = \exp(\tilde{x}^\top \tilde{\eta})$, where $\tilde{\eta} = (\eta_0, \eta_1, \dots, \eta_k)^\top$ denotes the corresponding vector of regression coefficients. Parameters can be estimated by the maximum-likelihood method, either directly maximizing the log-likelihood function or via a numerical method such as the Newton–Raphson procedure. The results obtained are reported in Table 6. The parameter estimates and related p-values calculated using the t test are displayed from left to right. Observe that the parameters α, β, constant η_0, and some of the covariates such as age, insure, countip, and countop are highly significant. The rest of the covariates do not seem to affect the response variable. At the standard nominal level, they are not significant. Note that as the parameter β is not located in the neighborhood of one, with the BGS model yielding a better fit to the data.

The coefficients for covariates in the model should be interpreted with caution, because the estimated coefficients are not the marginal effects. Thus, the marginal effects, for no dichotomous variables, are computed using the formula $\widehat{\eta}_k \exp(\tilde{x}^\top \widehat{\tilde{\eta}})$, where k represents the covariate k and $\widehat{\tilde{\eta}}$ is a vector of estimated parameters. These effects may be evaluated at the mean or at each observation. In our case, it can be seen that the age and countip variables have a positive effect on expenditure, while countop has a negative effect on it. For indicator

Table 7 Parameter estimates and p-values for the in-sample ABI claims data set including covariates

Variable	BSG		GB2	
	Estimate	p-value	Estimate	p-value
Attorney	1.249	< 0.01	1.196	< 0.01
Clmsex	− 0.037	0.709	− 0.021	0.825
Married	− 0.227	0.427	0.544	0.075
Single	− 0.499	0.091	0.347	0.270
Widowed	− 1.602	0.028	− 0.279	0.678
Clminsur	− 0.014	0.935	− 0.005	0.976
Seatbelt	− 0.803	< 0.01	− 0.565	0.097
Clmage	0.010	< 0.01	0.012	< 0.01
α or p	1.593	< 0.01	0.645	0.064
β or q	1.598	< 0.01	0.792	0.065
τ			0.523	0.015
Constant	1.702	< 0.01	− 0.020	0.970
NLL	1318.351		1320.366	
AIC	2658.701		2664.733	
BIC	2706.901		2717.315	
Observations	500		500	

variables such as insure, which takes only the value 0 or 1, the marginal effect in terms of the odds-ratio is approximately $\exp(\widehat{\eta}_j)$. Therefore, the conditional mean is $\exp(\widehat{\eta}_j)$ times greater if the indicator variable is one rather than zero. For the example considered, the dichotomous variable insure, going from not having insurance to having it, increases the logarithm of spending by approximately $\exp(0.093) \approx 1.09$, which represents a gross value of three monetary units.

4.4 In-sample and out-of-sample performance

Now, we examine the in-sample and out-of-sample performance of the BSG regression model. As a benchmark, we use the generalized beta case of the second kind (GB2) regression model given in Calderín-Ojeda et al. (2017). For that reason, we consider the Association of British Insurers' (ABI) claims data set, omitting the covariates that contain missing observations, resulting in a sample of 1,034 losses from a single state. We utilize the Claimant total economic loss as the response variable, and the factors indicated in Calderín-Ojeda et al. (2017) as covariates. We have randomly partitioned the ABI data set into two disjoint subsets of different sizes. The first subset of size 500 is used for fitting the models, whereas the second subset of size 534 tests the model's out-of-sample prediction accuracy. We have fitted the BSG and GB2 regression models to this in-sample data set, with the results being provided in Table 7. The parameter estimates and related p-values calculated using the t test are shown from left to right. Observe that the parameters α, β, constant η_0, and some of the covariates such as attorney, seatbelt, and clmage are highly significant. Note that for this particular sample, the fit of the BSG regression model is slightly better than the one provided by the GB2 regression model in terms of the three measures of model selection considered. Still, similar results are obtained for other samples of the same size.

Table 8 Null hypothesis, statistic, and decision for the Vuong test in the indicated set for ABI claims data

Set	Null hypothesis	Test statistic T	Decision at 5% significance level
In-sample	$\mathrm{E}[\ell_{\mathrm{BSG}}(\widehat{\boldsymbol{\theta}}_1) = \mathrm{E}[\ell_{\mathrm{GB2}}(\widehat{\boldsymbol{\theta}}_2)]$	-1.312	No rejection
Out-of-sample	$\mathrm{E}[\ell_{\mathrm{BSG}}(\widehat{\boldsymbol{\theta}}_1) = \mathrm{E}[\ell_{\mathrm{GB2}}(\widehat{\boldsymbol{\theta}}_2)]$	0.893	No rejection

Next, we evaluate these two regression models using the likelihood ratio test for non-nested models introduced by Vuong (1989). The Vuong test statistic is calculated as

$$T = \frac{1}{\omega\sqrt{n}}\left(\ell_f(\widehat{\boldsymbol{\theta}}_1) - \ell_g(\widehat{\boldsymbol{\theta}}_2) - \log\left(n\left(\frac{n_f}{2} - \frac{n_g}{2}\right)\right)\right),$$

where

$$\omega^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\log\left(\frac{f(y_i|\widehat{\boldsymbol{\theta}}_1)}{g(y_i|\widehat{\boldsymbol{\theta}}_2)}\right)\right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{f(y_i|\widehat{\boldsymbol{\theta}}_1)}{g(y_i|\widehat{\boldsymbol{\theta}}_2)}\right)\right)^2$$

is the sample variance of the pointwise log-likelihood ratios, with f and g representing the PDFs of two different non-nested models, $\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2$ are the maximum-likelihood estimates of $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$, and n_f, n_g are the number of estimated parameters in the model with PDFs $f(x|\boldsymbol{\theta}_1)$ and $g(x|\boldsymbol{\theta}_2)$, respectively. It is worth noting that the Vuong statistic is sensitive to the number of estimated parameters in each model, so that the test must be dimensionality corrected. We test H_0: $\mathrm{E}[\ell_f(\widehat{\boldsymbol{\theta}}_1) - \ell_g(\widehat{\boldsymbol{\theta}}_2)] = 0$ against H_0: $\mathrm{E}[\ell_f(\widehat{\boldsymbol{\theta}}_1) - \ell_g(\widehat{\boldsymbol{\theta}}_2)] \neq 0$. Here, T is asymptotically normally distributed under the null hypothesis. The rejection region for this test in favor of the alternative hypothesis occurs when $|T| > 1.96$ at the 5% significance level. Table 8 presents the null hypothesis, test statistic, and decision (at the 5% significance level for this comparison). For the ABI claim data set, the GB2 regression is not preferred to the BSG regression at 5%, both in-sample and out-of-sample.

5 Concluding remarks and future research

In this paper, the Birnbaum–Saunders gamma distribution was introduced. This novel probabilistic model was obtained by mixing the Birnbaum–Saunders and gamma distributions. After exploring some of its statistical properties, applications in credibility theory were examined. In particular, an analytical expression for the Bayes premium was provided, where the premium was written as a convex sum of a function of the individual data and the collective premium.

The model parameters were estimated via the maximum-likelihood method, and inference has been performed using this method to detect the significance of such parameters. Moreover, a simulation analysis was implemented to examine the behavior for the maximum-likelihood estimators and the correlation between these estimators.

The new distribution allowed the modeling of actuarial data associated with financial losses. Additionally, the model was proven to be helpful when calculating reinsurance premiums if the insurer adopts an excess-of-loss reinsurance agreement. As compared to other commonly used models in the actuarial literature, this new mixture distribution demonstrated a better fit to genuine insurance data across the whole empirical distribution. Moreover, a regression model was derived to explain the response variable, that is, claim size, as a function of a set of covariates providing a similar in-sample and out-of-sample to the generalized beta of the second kind of regression, which has traditionally been considered to explain heavy-tailed distributed data.

In summary, we can conclude that this distribution has two distinct advantages. On one hand, as a heavy-tailed model, the Birnbaum-Saunders gamma distribution is suitable for describing severity data. On the other hand, this distribution can be parameterized to incorporate factors and covariates that explain the mean of the response variable. The numerical evaluations of the suggested methodology allow us to demonstrate its high practical performance to characterize insurance data and prospective risk theory applications, making it helpful not only for actuaries but also for applied statisticians and data scientists.

Finally, some open problems that arose from this study are the following:

(i) The development of likelihood inferential methods by considering censored data, random effects, robustness, and reliability methods is of interest in these types of applications (see Villegas et al. 2011; Azevedo et al. 2012; Desousa et al. 2020).
(ii) Extensions to the multivariate case are also of practical relevance (see Marchant et al. 2016; Aykroyd et al. 2018; Sanchez et al. 2020).
(iii) Incorporation of time series, spatial and quantile regression structures in the modeling, as well as errors-in-variables, functional data analysis, and partial least squares regression, based on the proposed distribution, are also of interest (see Garcia-Papani et al. 2018; Martinez et al. 2019; Carrasco et al. 2020; Sanchez et al. 2020, 2021; Huerta et al. 2021; Korkmaz and Chesneau 2021; Leiva et al. 2021; Ribeiro et al. 2021; Figueroa-Zuniga et al. 2022; Saulo et al. 2022).
(iv) It is necessary to develop influence diagnostic tools to detect potentially influential cases, which are an important tool in all statistical modeling (see Liu et al. 2021; Sanchez et al. 2021).

Thus, our methodology encourages new challenges and provides inspiration for further research into other theoretical and numerical issues. These and other issues are being investigated, and the results should be communicated in a future work.

References

Albrecher H, Beirlant J, Eugels J (2017) Reinsurance: actuarial and statistical aspects. Wiley, New York
Arnold B (1983) Pareto distributions. International Cooperative Publishing House, Silver Spring
Aykroyd RG, Leiva V, Marchant C (2018) Multivariate Birnbaum–Saunders distributions: modelling and applications. Risks 6:21
Azevedo C, Leiva V, Athayde E, Balakrishnan N (2012) Shape and change point analyses of the Birnbaum–Saunders-t hazard rate and associated estimation. Comput Stat Data Anal 56:3887–3897
Beirlant J, Teugels J, Vynckier P (1996) Practical analysis of extreme values. Leuven University Press, Leuve
Beirlant J, Matthys GJ, Dierckx G (2005) Heavy-tailed distributions and rating. Astin Bull 31:41–62
Birnbaum ZW, Saunders SC (1969) A new family of life distributions. J Appl Probab 6:319–327
Boland P (2007) Statistical and probabilistic methods in actuarial science. Chapman and Hall, New York
Brooks C (2009) RATS handbook to accompany introductory econometrics for finance. Cambridge University Press, Cambridge
Bühlmann H, Gisler A (2005) A course in credibility theory and its applications. Springer, New York
Calderín-Ojeda E, Fergusson K, Wu X (2017) An EM algorithm for double-Pareto-lognormal generalized linear model applied to heavy-tailed insurance claims. Risks 5:60
Carrasco JMF, Figueroa-Zuniga J, Leiva V, Riquelme M, Aykroyd RG (2020) An errors-in-variables model based on the Birnbaum–Saunders and its diagnostics with an application to earthquake data. Stoch Environ Res Risk Assess 34:369–380
Desousa M, Saulo H, Leiva V, Santos-Neto M (2020) On a new mixture-based regression model: simulation and application to data with high censoring. J Stat Comput Simul 90:2861–2877

Efron B, Tibshirani R (1993) An introduction to the bootstrap. Chapman and Hall, New York
Embrechts P, Resnick S, Samorodnitsky G (1999) Extreme value theory as a risk management tool. N Am Actuarial J 3:30–41
Figueroa-Zuniga J, Bayes CL, Leiva V, Liu S (2022) Robust beta regression modeling with errors-in-variables: a Bayesian approach and numerical applications. Stat Pap. https://doi.org/10.1007/s00362-021-01260-1 (**in press**)
Garcia-Papani F, Leiva V, Uribe-Opazo MA, Aykroyd RG (2018) Birnbaum–Saunders spatial regression models: diagnostics and application to chemical data. Chemom Intell Lab Syst 177:114–128
Gómez-Déniz E (2008) A generalization of the credibility theory obtained by using the weighted balanced loss function. Insur Math Econ 42:850–854
Hashemi F, Naderi M, Jamalizadeh A (2019) Normal mean-variance Lindley Birnbaum-Saunders distribution. Stat Interface 12(4):585–597
Hashemi F, Naderi M, Mashinchi M (2019) Clustering right-skewed data stream via Birnbaum-Saunders mixture models: a flexible approach based on fuzzy clustering algorithm. Appl Soft Comput 82:105539
Huerta M, Leiva V, Liu S, Rodriguez M, Villegas D (2019) On a partial least squares regression model for asymmetric data with a chemical application in mining. Chemom Intell Lab Syst 190:55–68
Jessen AH, Mikosch T (2006) Regularly varying functions. Publ Inst Mat 80:171–192
Konstantinides D (2018) Risk theory. A heavy tail approach. World Scientific Publishing, New York
Korkmaz MÇ, Chesneau C (2021) On the unit Burr-XII distribution with the quantile regression modeling and applications. Comput Appl Math 40:29
Leiva V (2016) The Birnbaum–Saunders distribution. Academic Press, New York
Leiva V, Saulo H, Souza R, Aykroyd RG, Vila R (2021) A new BISARMA time series model for forecasting mortality using weather and particulate matter data. J Forecast 40:346–364
Liu S, Leiva V, Zhuang D, Ma T, Figueroa-Zuniga J (2021) Matrix differential calculus with applications in the multivariate linear model and its diagnostics. J Multivar Anal 188:104849
Marchant C, Leiva V, Cysneiros FJA (2016) A multivariate log-linear model for Birnbaum–Saunders distributions. IEEE Trans Reliab 65:816–827
Martinez S, Giraldo R, Leiva V (2019) Birnbaum–Saunders functional regression models for spatial data. Stoch Environ Res Risk Assess 33:1765–1780
Naderi M, Hashemi F, Bekker A, Jamalizadeh A (2020) Modeling right-skewed financial data streams: a likelihood inference based on the generalized Birnbaum–Saunders mixture model. Appl Math Comput 376:125109
Naderi M, Mozafari M, Okhli K (2020) Finite mixture modeling via skew-Laplace Birnbaum–Saunders distribution. J Stat Theory Appl 19:49–58
Ribeiro TF, Cordeiro GM, Peña-Ramírez FA, Guerra RR (2021) A new quantile regression for the COVID-19 mortality rates in the United States. Comput Appl Math 40(255):1–16
Rolski T, Schmidli H, Schmidt V, Teugel J (1999) Stochastic processes for insurance and finance. Wiley, New York
Ruskeepaa H (2009) Mathematica navigator. Mathematics, statistics, and graphics. Academic Press, New York
Sanchez L, Leiva V, Galea M, Saulo H (2020) Birnbaum-Saunders quantile regression models with application to spatial data. Mathematics 8:1000
Sanchez L, Leiva V, Galea M, Saulo H (2021) Birnbaum-Saunders quantile regression and its diagnostics with application to economic data. Appl Stoch Model Bus Ind 37:53–73
Saulo H, Dasilva A, Leiva V, Sanchez L, de la Fuente-Mella H (2022) Log-symmetric quantile regression models. Stat Neerl 76:124–163
Villegas C, Paula GA, Leiva V (2011) Birnbaum–Saunders mixed models for censored reliability data analysis. IEEE Trans Reliab 60:748–758
Vuong Q (1989) Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57:307–333
Wilcox R (2010) Fundamentals of modern statistical methods. Substantially improving power and accuracy. Springer, New York