Modeling right-skewed financial data streams: a likelihood inference based on the generalized Birnbaum-Saunders mixture model

Mehrdad Naderi[a,*], Farzane Hashemi[b], Andriette Bekker[a], Ahad Jamalizadeh[b,c]

[a] *Department of Statistics, Faculty of Natural & Agricultural Sciences, University of Pretoria, Pretoria, South Africa*
[b] *Department of Statistics, Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Kerman, Iran*
[c] *Mahani Mathematical Research Center, Shahid Bahonar University of Kerman, Kerman, Iran*

Abstract

Finite mixture models have recently been considered for analyzing positive support economical data streams with non-normal features. In this paper, a new mixture model based on the novel class of generalized Birnbaum-Saunders distributions is proposed to enhance strength and flexibility in modeling heterogeneous lifetime data. Some characteristics and properties of this mixture model are outlined. By presenting a convenient hierarchical representation, a mathematically elegant and computationally tractable EM-type algorithm is adopted for computing maximum likelihood estimates. Theoretical formulae of well-known risk measures referring to the class of generalized Birnbaum-Saunders distributions are derived. Finally, the utility of the postulated methodology is illustrated with some real-world data examples.

Keywords: Birnbaum-Saunders distribution, Finite mixture model, Normal mean-variance model, Risk measurement, Value-at-risk, Tail-Value-at-risk

1. Introduction

Asymmetric, leptokurtic and multimodal data are often presented in various fields such as insurance, econometrics, business, biology, genetics, industry and engineering. For modeling asymmetric and leptokurtic data, skewed and fat-tail distributions have commonly been used. For instance, [1] and [2] considered the skew-normal (SN) distribution [3] for evaluating some risk measures of insurance data. Eling [4, 5] exploited the SN and skew-*t* distributions [6] as promising benchmark models for describing the actuarial loss and asset returns of insurance companies. Shushi [7] identified the skew-elliptical distributions as alternative models in the context of risk measurement. Moreover, the generalized hyperbolic (GH) family of distributions has recently been considered for adequately modeling financial data. For example, Prause et al. [8] described the properties of the GH distribution and highlighted its practical features both through simulated and real datasets. Aas [9] demonstrated the superiority of two special cases of the GH distribution, namely the normal inverse Gaussian (NIG) and generalized hyperbolic skew-*t* (GHST) distributions, in evaluating Value-at-Risk (VaR) and Tail-Value-at-Risk (TVaR) measures. They showed the significance of the exponential/polynomial tail behavior in practice. More recently, Bee et al. [10] developed three extensions of the expectation-maximization (EM) algorithm for obtaining maximum likelihood estimate of the parameters of variance-gamma (VG) distribution. They showed that the VG model is flexible enough to accommodate skewness and leptokurtosis in order to model log-returns of financial assets.

The main reasons for using the aforementioned families of distributions in financial analyses are (1) the feasibility and potential for fast calibration of the parameters to data, and (2) the closeness of the family under combinations (the portfolio property). However, using the asymmetric distributions with $\mathbb{R}$ support is inappropriate in principle when the data are positive. More precisely, if the support is $\mathbb{R}^+$, using these distributions might cause boundary bias, that is,

*Corresponding author
Email addresses: m.naderi@up.ac.za (Mehrdad Naderi), farzane.hashemi1367@yahoo.com (Farzane Hashemi), andriette.bekker@up.ac.za (Andriette Bekker), a.jamalizadeh@uk.ac.ir (Ahad Jamalizadeh)

allocation of probability mass outside of the theoretical support. To overcome this deficiency, there has been a growing interest in seeking robust, positive and right-skewed distributions. See, for example, the work on the log-skew-normal [11], gamma [12–15], log-normal [14–16], inverse Gaussian [14, 17] and Weibull [18] distributions.

In recent decades, one of the most valuable lifetime models in applied statistics has been the two-parameter Birnbaum-Saunders (BS) distribution [19]. Properties and straightforward relationship with the normal distribution make the BS model a promising tool in engineering [20], medical sciences [21], wind analysis [22], as well as business and industry [23–26]. The BS distribution could be criticized not only for its lack of robustness against atypical observations (highly skewed and heavy-tailed data) but also for the fact that it cannot accommodate monotone (increasing or decreasing) nor bathtub-shaped hazard rate functions [27]. To overcome these deficiencies, some generalizations of the BS distribution have recently been proposed in [28–33]. Although these generalized models may not have physical meaning as the BS distribution, they can be used for modeling right-skewed and non-negative datasets with strong asymmetrical features. An interesting approach to extending the BS distribution, which is discussed in detail in the next section, is to replace the standard normal variable in the stochastic representation of the BS model with other random variables followed by highly skewed and heavy-tailed distributions. Based on this idea and exploiting normal mean-variance mixture (NMV) models, Hashemi et al. [34] introduced the normal mean-variance Lindley BS (NMVL-BS) distribution. They showed that the NMVL-BS distribution is robust in the presence of outliers and can model datasets with monotonic or non-monotonic hazard rate functions. Hashemi et al. [34] also concluded that the main advantage of the NMVL-BS distribution was its application to financial data. This advantage is due to its feasible fitting, its attractive properties, such as covering outliers, and its relationship with the NMV model.

Even though all the aforementioned distributions assume unimodality, it is generally accepted with the support of numerous empirical studies that economic and financial data streams often exhibit some heterogeneity, as reflected in multimodality, which is not compatible with the fitting of a single parametric distribution [35]. Realistic modeling thus calls for alternative statistical tools. The finite mixture (FM) model is a statistical tool aiming for classification and density estimation. Due to its flexibility and universal approximation capability with respect to complicated density functions in the presence of heterogeneity, the FM model has frequently been used in a broad variety of applications [21, 36–40].

This paper aims to (i) characterize a new generalization of the BS distribution by means of the NMV models, which we call the normal mean-variance generalized BS (NMV-GBS) distribution, (ii) propose the finite mixture of NMV-GBS (FM-NMV-GBS) distributions, (iii) illustrate the performance of the proposed model in dealing with financial datasets, and (iv) study two well-known risk measures, VaR and TVaR, for the proposed model. The remainder of the paper is therefore divided into five parts. The definition of the new model is presented in Section 2 and its parameter estimation procedure in Sections 3. In Section 4, we derive theoretical formulae for VaR and TVaR based on the FM-NMV-BS distributions. The advantages of the proposed model are examined in Section 5 through the analysis of four real financial datasets. Finally, Section 6 gives some concluding remarks and future considerations. Technical details and additional information are provided in the Online Supplement.

2. Normal mean-variance generalized Birnbaum-Saunders distribution

In order to introduce notation, we start with the definition of the NMV distribution. An in-depth discussion of the NMV model can be found in [41].

2.1. Preliminaries

2.1.1. The NMV model

A random variable X is said to follow the NMV model [42] if it has the stochastic representation

$$X = \mu + W\lambda + W^{1/2}\sigma Z, \tag{1}$$

where Z is distributed as a standard normal, $\mathcal{N}(0, 1)$, and independently of Z, W is a non-negative random variable with probability density function (pdf) $h(w; \boldsymbol{\theta})$. One can easily see that the conditional distribution of X given $W = w$ is $\mathcal{N}(\mu + w\lambda, w\sigma^2)$, and so the pdf of X can be expressed as

$$f_{\text{NMV}}(x; \mu, \lambda, \sigma^2, \boldsymbol{\theta}) = \int_0^\infty \frac{1}{\sqrt{2\pi w\sigma^2}} \exp\left\{-\frac{1}{2w\sigma^2}(x - \mu - w\lambda)^2\right\} h(w; \boldsymbol{\theta})\, dw, \qquad x \in \mathbb{R}.$$

In the rest of the paper, $f_{\text{NMV}}(x; \lambda, \boldsymbol{\theta})$ will represent the pdf of the NMV distribution with $\mu = 0$ and $\sigma = 1$, denoted by $X \sim \mathcal{NMV}(\lambda, \boldsymbol{\theta})$.

Depending on the chosen distribution for W in representation (1), the special cases of the NMV model, such as the NIG, GHST, VG, and skew-Laplace (SL) distributions, can be obtained. An important sub-class of the NMV model is the GH distribution. A random variable X is said to follow a GH distribution, denoted by $\mathcal{GH}(\mu, \sigma^2, \lambda, \kappa, \chi, \psi)$, if its pdf is

$$f_{\text{GH}}(x; \mu, \lambda, \sigma^2, \kappa, \chi, \psi) = C \frac{K_{0.5-\kappa}\left(\delta(x, \mu, \lambda, \sigma^2, \chi, \psi)\right)}{(\delta(x, \mu, \lambda, \sigma^2, \chi, \psi))^{0.5-\kappa}} \exp\left\{\frac{\lambda(x-\mu)}{\sigma^2}\right\}, \quad x \in \mathbb{R}, \tag{2}$$

where $\delta(x, \mu, \lambda, \sigma^2, \chi, \psi) = \sqrt{(\psi + (\lambda/\sigma)^2)(\chi + ((x-\mu)/\sigma)^2)}$, $C = \sqrt{(\psi/\chi)^{\kappa}}(\psi + (\lambda/\sigma)^2)^{0.5-\kappa} / \sqrt{2\pi\sigma^2} K_\kappa\left(\sqrt{\psi\chi}\right)$ is the normalizing constant and $K_a(\cdot)$ denotes the modified Bessel function of the third kind with order a. The notation $f_{\text{GH}}(\cdot; \lambda, \kappa, \chi, \psi)$ will be used to denote the pdf of standard GH distribution ($\mu = 0, \sigma = 1$). McNeil et al. [41] showed that X with pdf (2) can be generated by the NMV representation (1) when W has a generalized inverse Gaussian (GIG) [43] distribution, $W \sim \mathcal{GIG}(\kappa, \chi, \psi)$, with pdf

$$h_{\text{GIG}}(w; \kappa, \chi, \psi) = \left(\frac{\psi}{\chi}\right)^{\kappa/2} \frac{w^{\kappa-1}}{2K_\kappa\left(\sqrt{\psi\chi}\right)} \exp\left\{-\frac{1}{2}\left(w^{-1}\chi + w\psi\right)\right\}, \quad w > 0. \tag{3}$$

2.1.2. The BS distribution

The BS distribution is a positively skewed and unimodal distribution with non-negative support. An important property of the BS distribution is that it is closely related to the normal distribution by means of a simple stochastic representation. A random variable T is said to have a BS distribution with the shape and scale parameters α and β, respectively, if it can be expressed by

$$T = \frac{\beta}{4}\left[\alpha Z + \sqrt{(\alpha Z)^2 + 4}\right]^2, \tag{4}$$

where $Z \sim \mathcal{N}(0, 1)$. It can easily be shown that the cumulative distribution function (cdf) of T is

$$F(t; \alpha, \beta) = \Phi[c(t, \alpha, \beta)], \qquad t > 0,\ \alpha > 0,\ \beta > 0, \tag{5}$$

where $\Phi(\cdot)$ is the cdf of $\mathcal{N}(0, 1)$ and $c(t, \alpha, \beta) = \alpha^{-1}\left(\sqrt{t/\beta} - \sqrt{\beta/t}\right)$. The cdf of BS distribution can be formulated by a mixture of two GIG distributions with equal weight, i.e.

$$F(t; \alpha, \beta) = 0.5F_{\text{GIG}}\left(t; \frac{1}{2}, \frac{1}{\beta\alpha^2}, \frac{\beta}{\alpha^2}\right) + 0.5F_{\text{GIG}}\left(t; -\frac{1}{2}, \frac{1}{\beta\alpha^2}, \frac{\beta}{\alpha^2}\right).$$

An in-depth review and discussion of both the univariate and multivariate BS distributions can be found in [44–46], among others.

An appealing generalization of the BS distribution is obtained by replacing the standard normal variable Z in (4) with another random variable with a highly skewed and heavy-tailed distribution, or alternatively by replacing $\Phi(\cdot)$ in (5) with a cdf of asymmetric distribution. A new extension of the BS model is presented below by combining (1) and (4).

2.2. Model formulation

Definition 1. *Let T be a positive random variable. It follows the NMV-GBS distribution if it is related to the standard normal model via the following stochastic representation*

$$T = \frac{\beta}{4}\left[\alpha(W\lambda + W^{1/2}Z) + \sqrt{\left[\alpha(W\lambda + W^{1/2}Z)\right]^2 + 4}\right]^2, \tag{6}$$

where $Z \sim \mathcal{N}(0, 1)$ and $W \sim h(w; \boldsymbol{\theta})$ are independent. Consequently, the pdf of T is given by

$$f_{\text{NMV-GBS}}(t; \alpha, \beta, \lambda, \boldsymbol{\theta}) = C(t, \alpha, \beta) f_{\text{NMV}}(c(t, \alpha, \beta); \lambda, \boldsymbol{\theta}), \tag{7}$$

where $C(t, \alpha, \beta) = (t + \beta)/2\alpha\sqrt{\beta t^3}$ is the first derivative of $c(t, \alpha, \beta)$.

Denote $T \sim \mathcal{NMV}-\mathcal{GBS}(\alpha,\beta,\lambda,\boldsymbol{\theta})$ for short, if T has pdf (7). Note that the parameter σ^2 of the NMV distribution is fixed to avoid problem of identifiability. The NMV-GBS quantile function is given by

$$Q(u) = \frac{\beta}{4}\left[\alpha F^{-1}_{\text{NMV}}(u;\lambda,\boldsymbol{\theta}) + \sqrt{\left(\alpha F^{-1}_{\text{NMV}}(u;\lambda,\boldsymbol{\theta})\right)^2 + 4}\right]^2, \quad u \in (0,1),$$

where $F^{-1}_{\text{NMV}}(\cdot;\lambda,\boldsymbol{\theta})$ denotes the quantile function of the $\mathcal{NMV}(\lambda,\boldsymbol{\theta})$ model.

Proposition 1. *The NMV-GBS distribution is a member of the scale family of distributions and is closed under reciprocation. i.e. if $T \sim \mathcal{NMV}-\mathcal{GBS}(\alpha,\beta,\lambda,\boldsymbol{\theta})$, then $aT \sim \mathcal{NMV}-\mathcal{GBS}(\alpha,a\beta,\lambda,\boldsymbol{\theta})$ for any $a > 0$, and $T^{-1} \sim \mathcal{NMV}-\mathcal{GBS}(\alpha,\beta^{-1},\lambda,\boldsymbol{\theta})$.*

Proof. The proof of proposition can be found in Appendix *B* of the Online Supplement. □

Proposition 2. *Let $T \sim \mathcal{NMV}-\mathcal{GBS}(\alpha,\beta,\lambda,\boldsymbol{\theta})$. Then, the conditional distribution of T given by $W = w$ is*

$$T|W = w \sim \mathcal{EBS}(\alpha\sqrt{w},\beta,2,-\sqrt{w}\lambda,0),$$

where $\mathcal{EBS}$ denotes the extended BS distribution [47]. Thus, the pdf of $T|W = w$ is given by

$$f_{T|W=w}(t;\alpha,\beta,\lambda) = \frac{C(t,\alpha,\beta)}{\sqrt{2\pi w}}\exp\Big\{-\frac{1}{2w}(c(t,\alpha,\beta)-\lambda w)^2\Big\}.$$

Proof. The proof of proposition can be found in Appendix *B* of the Online Supplement. □

Assume in (6), $W \sim \mathcal{GIG}(\kappa,\chi,\psi)$. Then, an interesting sub-class of the NMV-GBS model called the generalized hyperbolic BS distribution (GH-BS), denoted by $\mathcal{GH}-\mathcal{BS}(\alpha,\beta,\lambda,\kappa,\chi,\psi)$, is obtained. The following propositions provide limiting and special cases of the GH-BS distribution.

Proposition 3. *Let $T \sim \mathcal{GH}-\mathcal{BS}(\alpha,\beta,\lambda,\kappa,\chi,\psi)$.*

i. If ψ approaches zero and $\kappa = -\nu/2, \chi = \nu$, then T tends to the GHST-BS distribution. Here, Student-t BS (T-BS) model is also obtained if λ tends to zero.
ii. By setting $\kappa = 1$ and $\psi = 0$, the hyperbolic BS (H-BS) and variance-gamma BS (VG-BS) distributions follow, respectively.
iii. When $\kappa = -0.5$, the GIG model becomes the inverse Gaussian distribution, and its corresponding GH-BS distribution is the normal inverse Gaussian BS (NIG-BS) model.
iv. If $\kappa = 1, \psi = 1, \chi = 0$, the GIG model becomes the exponential distribution and its corresponding GH-BS distribution is the skew-Laplace BS (SL-BS) model [39]. Here, the Laplace BS (L-BS) distribution can be obtained if λ approaches zero.
v. The GH-BS distribution includes the scale-mixture Birnbaum-Saunders distribution [33] as λ tends to zero.

Proof. The proof of proposition can be found in Appendix *B* of the Online Supplement. □

Proposition 4.

i. The original BS distribution is obtained from (6) when $h(\cdot;\boldsymbol{\theta})$ degenerates to 1 and λ tends to zero.
ii. Let W in (6) be distributed as $\mathcal{BS}(\tau,1)$. Then, the pdf of random variable T followed by the NMVBS-BS distribution can be presented by

$$f_{\text{NMVBS-BS}}(t;\alpha,\beta,\lambda,\tau) = 0.5C(t,\alpha,\beta)\Big(f_{GH}(c(t,\alpha,\beta);\lambda,0.5,\tau^{-2},\tau^{-2}) + f_{GH}(c(t,\alpha,\beta);\lambda,-0.5,\tau^{-2},\tau^{-2})\Big).$$

Details of the NMVBS distribution can be found in [48, 49].

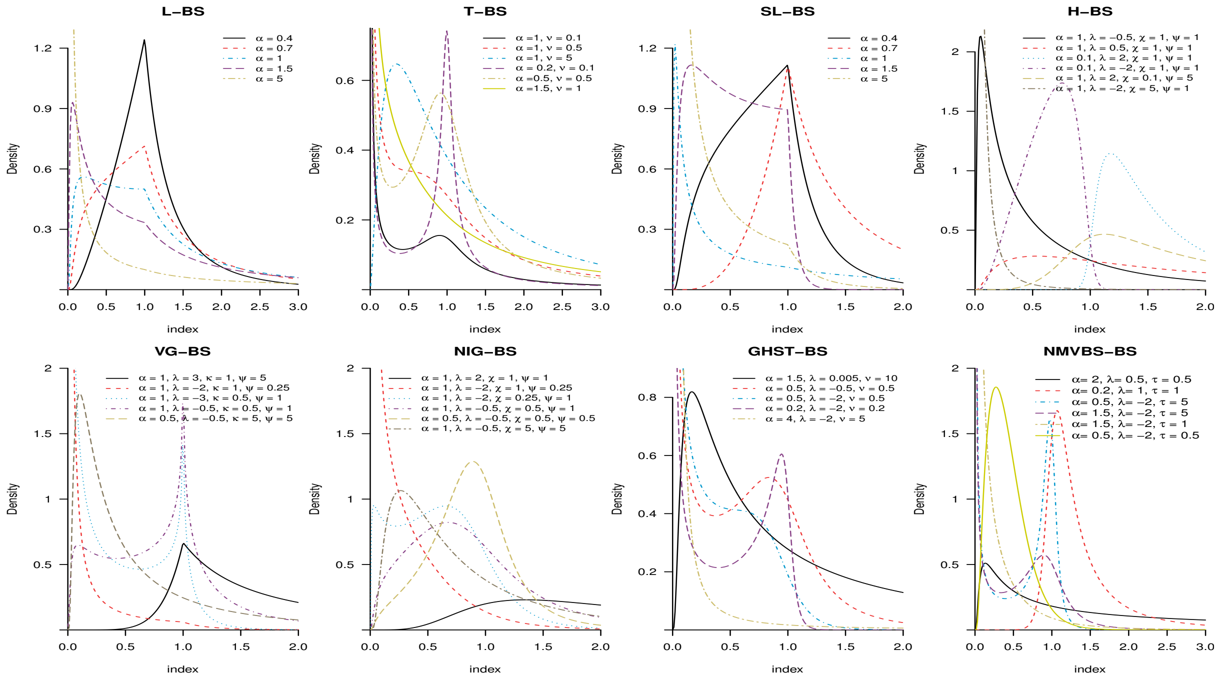

Figure 1: Density curves of special cases of the GH-BS distribution for arbitrary parameter choices.

iii. *Let W in (6) be distributed as Lindley distribution with shape parameter τ, denoted by $\mathcal{L}(\tau)$. Then, the random variable T follows the NMVL-BS distribution with the following pdf*

$$f_{\text{NMVL-BS}}(t;\alpha,\beta,\lambda,\tau) = \frac{C(t,\alpha,\beta)}{1+\tau}\Big(\tau f_{GH}(c(t,\alpha,\beta);\lambda,1,0,2\tau) + f_{GH}(c(t,\alpha,\beta);\lambda,2,0,2\tau)\Big).$$

Details of the NMVL and NMVL-BS distributions can be found in [34, 50].

Proof. The proof of proposition can be found in Appendix *B* of the Online Supplement. □

The following theorem is crucial for calculating some of the conditional expectations involved in the proposed EM-type algorithm in the next section.

Theorem 1. *Let $W \sim \mathcal{GIG}(\kappa,\chi,\psi)$ and $T \sim \mathcal{GH}-\mathcal{BS}(\alpha,\beta,\lambda,\kappa,\chi,\psi)$. Then, the conditional distribution of W given $T = t$ is $\mathcal{GIG}(\kappa - 0.5, \chi+\rho, \psi+\lambda^2)$, where $\rho = c^2(t,\alpha,\beta)$. Moreover,*

$$E[W^r|T=t] = \Big(\frac{\chi+\rho}{\psi+\lambda^2}\Big)^{r/2} R_{(\kappa,r)}\left(\sqrt{(\chi+\rho)(\psi+\lambda^2)}\right), \qquad \textit{for} \quad r = \pm1, \pm2, \ldots,$$

$$E[\log W|T=t] = \frac{\partial E[W^{\vartheta}|T=t]}{\partial\vartheta}\Big|_{\vartheta=0} = \log\left(\sqrt{\frac{\chi+\rho}{\psi+\lambda^2}}\right) + \frac{1}{K_{\kappa-0.5}(\sqrt{(\chi+\rho)(\psi+\lambda^2)})}\frac{\partial}{\partial\kappa}K_{\kappa-0.5}(\sqrt{(\chi+\rho)(\psi+\lambda^2)}),$$

where $R_{(\kappa,r)}(a) = K_{\kappa+r}(a)/K_{\kappa}(a)$.

Proof. The proof of theorem is provided in Appendix *B* of the Online Supplement. □

Table 1 summarizes the conditional distribution of W given $T = t$ for the limiting and special cases described in Propositions 3 and 4.

3. Finite mixture of the NMV-GBS distributions

Let $T_1, \ldots, T_n$ be n independent random variables taken from a FM-NMV-GBS distributions. The pdf of a g-component FM-NMV-GBS model is given by

$$f(t_j;\boldsymbol{\Theta}) = \sum_{i=1}^{g} \pi_i f_{\text{NMV-GBS}}(t_j;\alpha_i,\beta_i,\lambda_i,\boldsymbol{\theta}_i), \tag{8}$$

where π_i is a mixing proportion of the ith sub-population that is constrained to be $\pi_i > 0$ with the constraint $\sum_{i=1}^{g}\pi_i = 1$, $\boldsymbol{\Psi}_i = (\alpha_i,\beta_i,\lambda_i,\boldsymbol{\theta}_i)$ and $\boldsymbol{\Theta} = (\pi_1,\ldots,\pi_{g-1},\boldsymbol{\Psi}_1,\ldots,\boldsymbol{\Psi}_g)$. Therefore, the log-likelihood function of $\boldsymbol{\Theta}$ associated with the observed data $\boldsymbol{t} = (t_1,\ldots,t_n)^\top$ can be obtained as

$$\ell(\boldsymbol{\Theta}|\boldsymbol{t}) = \sum_{j=1}^{n} \log\left(\sum_{i=1}^{g} \pi_i f_{\text{NMV-GBS}}(t_j;\boldsymbol{\Psi}_i)\right). \tag{9}$$

To determine the ML estimate of the parameters in (9), a direct maximization is tortuous. This difficulty in obtaining parameter estimates is due to the complicated derivatives with respect to the parameters. An alternative framework for computing the ML estimate is the expectation-maximization (EM) algorithm [51]. The EM algorithm is an iterative procedure of parameter estimation, that was originally used with incomplete data. To apply the EM-type algorithm to finite mixture models, it is convenient to introduce a set of missing component membership labels $\boldsymbol{Z}_j = (Z_{1j},\ldots,Z_{gj})^\top$, where $Z_{ij} = 1$ if observation j is in component i and $Z_{ij} = 0$ otherwise, for $j = 1,\ldots,n$, and $i = 1,\ldots,g$. This implies that $\boldsymbol{Z}_j$ independently follows a multinomial distribution with one trial and probabilities $(\pi_1,\ldots,\pi_g)$, denoted by $\boldsymbol{Z}_j \sim \mathcal{M}(1;\pi_1,\ldots,\pi_g)$. This leads to presenting the hierarchical formulation of (8) by Proposition 2 as

$$\begin{aligned} T_j|(W = w_j, Z_{ij} = 1) &\sim \mathcal{EBS}(\alpha_i\sqrt{w_j},\beta_i,2,-\lambda_i\sqrt{w_j}),\\ W_j|Z_{ij} = 1 &\sim h(w_j;\boldsymbol{\theta}_i),\\ \boldsymbol{Z}_j &\sim \mathcal{M}(1;\pi_1,\ldots,\pi_g). \end{aligned}$$

Therefore, the complete-data log-likelihood function of $\boldsymbol{\Theta}$ given the observed data $\boldsymbol{t}$ and hidden variables $\boldsymbol{w} = (w_1,\ldots,w_n)^\top$ and $\boldsymbol{Z} = (\boldsymbol{Z}_1,\ldots,\boldsymbol{Z}_n)^\top$, omitting additive constants, is

$$\ell_c(\boldsymbol{\Theta}|\boldsymbol{t},\boldsymbol{w},\boldsymbol{Z}) = \sum_{j=1}^{n}\sum_{i=1}^{g} z_{ij}\left\{\log\pi_i + \log h(w_j;\boldsymbol{\theta}_i) - \log\alpha_i - \frac{1}{2}\log\beta_i + \log(t_j+\beta_i) - \frac{w_j^{-1}}{2\alpha_i^2}S_j - \frac{\lambda_i^2}{2}w_j + \frac{\lambda_i}{\alpha_i}a(t_j,1,\beta_i)\right\}, \tag{10}$$

where $S_j = (t_j/\beta_i - \beta_i/t_j - 2)$.

3.1. Parameter estimation via ECME algorithm

In this section, the Expectation Conditional Maximization Either (ECME) algorithm is exploited to compute the parameters estimate of the FM-NMV-GBS model. The ECME algorithm was originally proposed by [52] as an extension of the ECM algorithm [53]. The ECME algorithm is implemented by replacing the maximization (M) step of the EM algorithm with a sequence of computationally simpler conditional maximization (CM) steps in which, as explained below, the so-called Q-function or the corresponding constrained actual likelihood function is maximized.

The ECME algorithm for obtaining the ML estimate of the FM-NMV-GBS distributions proceeds as follows.

At the iteration k, we compute the so-called Q-function, in the E-step, which is defined as the expected value of complete-data log-likelihood (10) with $\boldsymbol{\Theta}$ valued at $\hat{\boldsymbol{\Theta}}^{(k)}$

$$Q(\boldsymbol{\Theta}|\hat{\boldsymbol{\Theta}}^{(k)}) = \sum_{j=1}^{n}\sum_{i=1}^{g}\hat{z}_{ij}^{(k)}\left\{\log\pi_i - \log\alpha_i - \frac{1}{2}\log\beta_i + \log(t_j+\beta_i) - \frac{\hat{u}_{ij}^{(k)}}{2\alpha_i^2}S_j - \frac{\lambda_i^2}{2}\hat{w}_{ij}^{(k)} + \frac{\lambda_i}{\alpha_i}a(t_j,1,\beta_i)\right\} + \sum_{j=1}^{n}\sum_{i=1}^{g}\hat{z}_{ij}^{(k)}\hat{\Upsilon}_{ij},$$

$$(11)$$

where $\hat{u}_{ij}^{(k)} = E[W_j^{-1}|t_j, Z_{ij} = 1, \hat{\mathbf{\Theta}}_i^{(k)}]$, $\hat{w}_{ij}^{(k)} = E[W_j|t_j, Z_{ij} = 1, \hat{\mathbf{\Theta}}_i^{(k)}]$ and $\hat{\Upsilon}_{ij} = E[\log h(W_j; \boldsymbol{\theta}_i)|t_j, Z_{ij} = 1, \hat{\mathbf{\Theta}}_i^{(k)}]$ are calculated by Theorem 1 and $\hat{z}_{ij} = E[Z_{ij}|t_j, \hat{\mathbf{\Theta}}] = f_{\text{NMV-GBS}}(t_j; \alpha_i, \beta_i, \lambda_i, \boldsymbol{\theta}_i)/f(t_j; \mathbf{\Theta})$.

Let $n_i = \sum_{j=1}^n \hat{z}_{ij}^{(k)}$, $A_i = \sum_{j=1}^n \hat{z}_{ij}^{(k)}\hat{w}_{ij}^{(k)}$ and $B_i = \sum_{j=1}^n \hat{z}_{ij}^{(k)}\hat{u}_{ij}^{(k)}$, and update $\hat{\mathbf{\Theta}}^{(k)}$ by maximizing (11) over $\mathbf{\Theta}$. This leads to the following CM estimators:

CM-step 1. Update the parameters $\hat{\pi}_i^{(k)}, \hat{\alpha}_i^{(k)}, \hat{\beta}_i^{(k)}$, and $\hat{\lambda}_i^{(k)}$ as

$$\begin{aligned}
\hat{\pi}_i^{(k+1)} &= \frac{n_i}{n}, \qquad \hat{\lambda}_i^{(k+1)} = \frac{1}{A_i}\sum_{j=1}^n \hat{z}_{ij}^{(k)} a(t_j; \hat{\alpha}_i^{(k)}, \hat{\beta}_i^{(k)}),\\
\hat{\alpha}_i^{2(k+1)} &= \frac{1}{n_i\hat{\beta}_i^{(k)}}\sum_{j=1}^n \hat{z}_{ij}^{(k)}\hat{u}_{ij}^{(k)} t_j + \frac{\hat{\beta}_i^{(k)}}{n_i}\sum_{j=1}^n \hat{z}_{ij}^{(k)}\frac{\hat{w}_{ij}^{(k)}}{t_j} - 2\frac{B_i}{n_i} - \frac{1}{A_i n_i}\Big(\sum_{j=1}^n \hat{z}_{ij}^{(k)} a(t_j, 1, \hat{\beta}_i^{(k)})\Big)^2,\\
\hat{\beta}_i^{(k+1)} &= \arg\max_{\beta\in\mathbb{R}^+} \ell_{obs}(\beta_i; \hat{\alpha}_i^{(k+1)}, \hat{\lambda}_i^{(k+1)}),
\end{aligned}$$

where

$$\ell_{obs}(\beta_i; \hat{\alpha}_i^{(k+1)}, \hat{\lambda}_i^{(k+1)}) = \sum_{j=1}^n \hat{z}_{ij}^{(k)}\Big\{-\frac{1}{2}\log\beta_i - \frac{\hat{u}_{ij}^{(k)}}{2\hat{\alpha}_i^{2(k+1)}}S_j + \log(t_j + \beta_i) - \frac{\hat{\lambda}_i^{2(k+1)}}{2}\hat{w}_{ij}^{(k)} + \frac{\hat{\lambda}_i^{(k+1)}}{\hat{\alpha}_i^{(k+1)}}a(t_j, 1, \beta_i)\Big\}.$$

CM-step 2. The update of $\hat{\boldsymbol{\theta}}^{(k)}$ is obtained by maximizing (11), or alternatively $\sum_{j=1}^n\sum_{i=1}^g \hat{z}_{ij}^{(k)}\hat{\Upsilon}_{ij}$. Table 1 includes some special cases of the FM-NMV-GBS distributions along with the closed form of $\hat{\Upsilon}_{ij}$ and their corresponding update of $\hat{\boldsymbol{\theta}}^{(k)}$.

3.2. *Computational aspects*

To implement the proposed ECME algorithm, we recommend the following initialization, convergence rule and model selection criteria.

3.2.1. *Initialization*

Generating admissible starting values is crucial to achieving swift convergence. Moreover, since the EM algorithm may not give a maximum global solution if the initial value is too far from the real parameter value, the choice of a good starting point plays an important role in parameter estimation. The adopted strategy for obtaining reasonable starting values is summarized as follows:

1. Separate the sample into g groups using the k-means cluster algorithm via R function "`kmeans( )`".
2. Compute the proportion of data points belonging to the same cluster i, and use them as an initial value of π_i.
3. For each group, create the initial value $\alpha_i^{(0)}$ and $\beta_i^{(0)}$, for example, by using the modified moment estimates proposed by [54], and $\lambda_i^{(0)}$ the skewness of ith group.
4. The initial value of $\boldsymbol{\theta}_i$ can be obtained as a moment or ML estimate of the baseline distribution on W for each group.

3.2.2. *Stopping rule*

To determine convergence of the EM algorithm, we use the Aitken acceleration method [55] in order to evade an indication of lack of progress of the algorithm [56]. At iteration $k+1$, the asymptotic estimate of the log-likelihood [57] is

$$\ell_\infty(\hat{\boldsymbol{\theta}}^{(k+1)}) = \ell(\hat{\boldsymbol{\theta}}^{(k+1)}) + \frac{1}{1-a^{(k)}}\Big\{\ell(\hat{\boldsymbol{\theta}}^{(k+1)}) - \ell(\hat{\boldsymbol{\theta}}^{(k)})\Big\},$$

where the Aitken acceleration factor is calculated as $a^{(k)} = (\ell(\hat{\boldsymbol{\theta}}^{(k+1)}) - \ell(\hat{\boldsymbol{\theta}}^{(k)}))/(\ell(\hat{\boldsymbol{\theta}}^{(k)}) - \ell(\hat{\boldsymbol{\theta}}^{(k-1)}))$. Therefore, the algorithm can be considered to have reached convergence if $\ell_\infty(\hat{\boldsymbol{\theta}}^{(k+1)}) - \ell(\hat{\boldsymbol{\theta}}^{(k)}) < \varepsilon$ [58]. In our study, the tolerance ε is set equal to 10^{-5}.

Table 1: Special cases of the FM-NMV-GBS distributions along with their mixing corresponding parameters estimate.

FM-NMV-GBS sub-model	Mixing parameter ($\boldsymbol{\Psi}_i$)	Mixing distribution (W_j)	Conditional distribution ($W_j \mid T = t_j, Z_{ij} = 1$)	$\hat{\Upsilon}_{ij}$ (omitted additive constants)	Estimator of $\boldsymbol{\theta}_i$
L-BS	(α_i, β_i)	$\mathcal{GIG}(1, 0, 1)$	$\mathcal{GIG}(0.5, \rho_{ij}, 1)$	———	No parameter to estimate
T-BS	$(\alpha_i, \beta_i, \nu_i)$	$\mathcal{GIG}(-\nu_i/2, \nu_i, 0)$	$\mathcal{GIG}(-(\nu_i + 1)/2, \nu_i + \rho_{ij}, 0)$	$\frac{\nu_i}{2}(\log \frac{\nu_i}{2} - \hat{w}_{ij} - \hat{u}_{ij}) - \log \Gamma(\frac{\nu_i}{2})$	$\hat{\nu}_i$s are obtained numerically
SL-BS	$(\alpha_i, \beta_i, \lambda_i)$	$\mathcal{GIG}(1, 0, 1)$	$\mathcal{GIG}(0.5, \rho_{ij}, 1 + \lambda_i^2)$	———	No parameter to estimate
H-BS	$(\alpha_i, \beta_i, \lambda_i, \chi_i, \psi_i)$	$\mathcal{GIG}(1, \chi_i, \psi_i)$	$\mathcal{GIG}(0.5, \chi_i + \rho_{ij}, \psi_i + \lambda_i^2)$	$0.5 \log \frac{\psi_i}{\chi_i} - \log K_1(\sqrt{\chi_i \psi_i})$ $-0.5(\chi_i \hat{u}_{ij} + \psi_i \hat{w}_{ij})$	$(\hat{\chi}_i, \hat{\psi}_i)$s are obtained numerically
VG-BS	$(\alpha_i, \beta_i, \lambda_i, \kappa_i, \psi_i)$	$\mathcal{GIG}(\kappa_i, 0, \psi_i)$	$\mathcal{GIG}(\kappa_i\text{-}0.5, \rho_{ij}, \psi_i + \lambda_i^2)$	$\kappa_i \log \psi_i - \log \Gamma(\kappa_i) - \kappa_i \log 2$ $-\hat{w}_{ij}\psi_i/2 + (\kappa - 1)\widehat{Lw}_{ij}$ where $\widehat{Lw}_{ij} = E[\log W \mid T = t_j, \boldsymbol{\theta}_i]^\star$	$(\hat{\kappa}_i, \hat{\psi}_i)$s are obtained numerically
NIG-BS	$(\alpha_i, \beta_i, \lambda_i, \chi_i, \psi_i)$	$\mathcal{GIG}(-0.5, \chi_i, \psi_i)$	$\mathcal{GIG}(-1, \chi_i + \rho_{ij}, \psi_i + \lambda_i^2)$	$0.5(\log \chi_i - \hat{u}_{ij}\chi_i - \hat{w}_{ij}\psi_i) + \sqrt{\chi_i \psi_i}$	$\hat{\chi}_i = \upsilon_i A_i/n_i$, and $\hat{\psi}_i = \upsilon_i^2/\psi_i$ where $\upsilon_i = -(1 - \frac{A_i B_i}{n_i^2})^{-1}$
GHST-BS	$(\alpha_i, \beta_i, \lambda_i, \nu_i)$	$\mathcal{GIG}(-\nu_i/2, \nu_i, 0)$	$\mathcal{GIG}(-(\nu_i + 1)/2, \nu_i + \rho_{ij}, \lambda_i^2)$	$\frac{\nu_i}{2}(\log \frac{\nu_i}{2} - \hat{w}_{ij} - \hat{u}_{ij}) - \log \Gamma(\frac{\nu_i}{2})$	$\hat{\nu}_i$s are obtained numerically
NMVBS-BS	$(\alpha_i, \beta_i, \lambda_i, \tau_i)$	$\mathcal{BS}(\tau_i, 1)$	Mixture of two GIG distributions: $\mathcal{GIG}(0, \tau_i^{-2} + \rho_{ij}, \tau_i^{-2} + \lambda_i^2)$ $\mathcal{GIG}(-1, \tau_i^{-2} + \rho_{ij}, \tau_i^{-2} + \lambda_i^2)$ with the weights $P_{1i}(t_j)$	$-\log \tau_i + \frac{\hat{w}_{ij} + \hat{u}_{ij} - 2}{2\tau_i^2}$	$\hat{\tau}_i = \sqrt{\sum_{j=1}^{n_i} \hat{z}_{ij}(\hat{w}_{ij} + \hat{u}_{ij} - 2)/n_i}$
NMVL-BS	$(\alpha_i, \beta_i, \lambda_i, \tau_i)$	$\mathcal{L}(\tau_i)$	Mixture of two GIG distributions: $\mathcal{GIG}(0.5, \rho_{ij}, 2\tau_i + \lambda_i^2)$ $\mathcal{GIG}(1.5, \rho_{ij}, 2\tau_i + \lambda_i^2)$ with the weights $P_{2i}(t_j)$	$\log \frac{\tau_i^2}{1 + \tau_i} - \tau_i \hat{w}_{ij}$	$\hat{\tau}_i = \frac{(1 - \bar{w}_i) + \sqrt{(1 - \bar{w}_i) + 8\bar{w}_i}}{2\bar{w}_i}$ where $\bar{w}_i = A_i/n_i$

Notation: $\rho_{ij} = c^2(t_j, \alpha_i, \beta_j)$, $P_{1i}(t_j) = f_{GH}(\rho_{ij}; 0, \lambda_i, 1, 0.5, \tau_i^{-2}, \tau_i^{-2})/[f_{GH}(\rho_{ij}; 0, \lambda_i, 1, 0.5, \tau_i^{-2}, \tau_i^{-2}) + f_{GH}(\rho_{ij}; 0, \lambda_i, 1, -0.5, \tau_i^{-2}, \tau_i^{-2})]$, and $P_{2i}(t_j) = \tau_i f_{GH}(\rho_{ij}; 0, \lambda_i, 1, 1, 0, 2\tau_i)/(\tau_i f_{GH}(\rho_{ij}; 0, \lambda_i, 1, 1, 0, 2\tau_i) + f_{GH}(\rho_{ij}; 0, \lambda_i, 1, 2, 0, 2\tau_i))$.

* The expectation $\widehat{Lw}_{ij}$ can be obtained by the "`Egig( )`" function in the `R` package "**ghyp**".

3.2.3. Model selection and performance evaluation

For the sake of comparison, various information criteria that take the form of a penalized log-likelihood $mC(n) - 2\ell_{\max}$ were introduced for model selection, where $\ell_{\max}$ is the maximized log-likelihood and m is the number of free parameters in the considered model. In our data analysis, the Bayesian information criterion (BIC) [59] with the penalty term $C(n) = \log n$ is adopted. Note that the model with the smallest BIC value is selected to fit the data.

Remark 1. *If the GH-BS distribution is considered as a mixing component in the FM-NMV-GBS model, an elegant reparametrization, known as $\bar{\alpha}$-parametrization, can be used to reduce the number of free parameters. Let $\omega = R_{(\kappa,1)}(\sqrt{\chi\psi})\sqrt{\chi/\psi}$. Then, the following formulae can be used to switch between the parameterization $(\alpha, \beta, \lambda, \kappa, \chi, \psi)$ and $(\alpha, \beta, \lambda^*, \kappa, \bar{\alpha})$:*

$$\bar{\alpha} = \sqrt{\chi\psi} \quad \text{and} \quad \lambda^* = \omega\lambda.$$

Clearly, in the $\bar{\alpha}$-parametrization of GH-BS distribution, the scale parameter of GH model ($\sigma^2 = 1$) changes to $\sigma^2 = \omega$. This condition may leads to the identifiability issue if $\bar{\alpha}$-parametrization is used in parameter estimation. Another drawback of $\bar{\alpha}$-parametrization is that it does not exist when $\bar{\alpha}$ approaches zero. In our data analysis, we use $\bar{\alpha}$-parametrization for the FM-H-BS and FM-NIG-BS distributions since they do not pose identifiability challenges.

4. Risk measure for the FM-NMV-GBS distributions

Risk evaluation is important to investors who hold portfolios of risky assets. Risk measures and associated theories thus play important roles in estimating financial losses. Among the several purposes of risk measures, the most important ones in practice are: determination of risk capital and capital adequacy, management tools and insurance premiums [41]. To attain these purposes, statistical tools play a substantial role since most modern measures of risk in a portfolio are statistical quantities. Based on the following proposition, we can calculate different risk measures on the assumption that the asset features are followed by the NMV-GBS distribution.

Theorem 2. *Let $W \sim \mathcal{GIG}(\kappa, \chi, \psi)$. Then,*

$$E(W^r \phi(t; \lambda W, W)) = \frac{\left(\frac{\psi}{\chi}\right)^{\kappa/2} K_{\kappa+r-0.5}\left(\sqrt{(\psi+\lambda^2)(\chi+t^2)}\right)}{K_\kappa(\sqrt{\psi\chi})\sqrt{2\pi}} \left(\frac{\chi+t^2}{\psi+\lambda^2}\right)^{(\kappa+r-0.5)/2} \exp\{t\lambda\}, \qquad r \in \mathbb{R}. \tag{12}$$

Proof. The proof follows from the pdfs of normal and GIG distributions. Using (3), we have

$$\begin{aligned} E(W^r \phi(t; \lambda W, W)) &= \frac{\exp\{t\lambda\}}{\sqrt{2\pi} K_\kappa(\sqrt{\psi\chi})} \left(\frac{\psi}{\chi}\right)^{\kappa/2} \int_0^\infty \frac{w^{\kappa+r-0.5-1}}{2} \exp\left\{\frac{-1}{2}\left(w^{-1}(\chi+t^2) + w(\psi+\lambda^2)\right)\right\} dw \\ &= \frac{\left(\frac{\psi}{\chi}\right)^{\kappa/2} K_{\kappa+r-0.5}\left(\sqrt{(\psi+\lambda^2)(\chi+t^2)}\right)}{K_\kappa(\sqrt{\psi\chi})\sqrt{2\pi}} \left(\frac{\chi+t^2}{\psi+\lambda^2}\right)^{(\kappa+r-0.5)/2} \exp\{t\lambda\}. \end{aligned}$$

□

Remark 2. *Based on Table 1 and Theorem 2, we can obtain a closed form of (12) for the special cases of the NMV-GBS distribution.*

The risk measures considered here are: the target shortfall (TS), VaR, and more importantly, TVaR measure. For the sake of notation, we define the rth order upper partial moment of the random variable and the probability of shortfall (PS) and outperformance (PO) as follows:

Let $\mathbb{E}[T - t_q]_+^r$ be the rth order upper partial moment of the random variable T with respect to $t_q \in \mathbb{R}^+$. More specifically,

$$\mathbb{E}[T - t_q]_+^r = \int_{t_q}^\infty (t - t_q)^r g_T(t; \boldsymbol{\omega})\, dt,$$

where t_q is a target separating gains and losses, and $f_T(\cdot; \boldsymbol{\omega})$ is the pdf of T parameterized with $\boldsymbol{\omega}$. The reference point t_q can be specified as a fixed target, e.g. a given income poverty line that applies to all households equally, or as a moving target, i.e. the target is not fixed but depends on the household-specific distribution of the random variable [60]. In the following proposition, we provide explicit formulae for evaluating the PS and PO for the NMV-GBS model.

Proposition 5. *Let $T \sim \mathcal{NMV}-\mathcal{GBS}(\alpha, \beta, \lambda, \boldsymbol{\theta})$. The probability of T that falls short or outperforms a target level t_q can be obtained, respectively, by*

$$PS(t_q, \alpha, \beta, \lambda, \boldsymbol{\theta}) = 1 - \mathbb{E}[T - t_q]_+^0 = F_{NMV}(c(t_q, \alpha, \beta); \lambda, \boldsymbol{\theta}),$$

and

$$PO(t_q, \alpha, \beta, \lambda, \boldsymbol{\theta}) = \mathbb{E}[T - t_q]_+^0 = 1 - F_{NMV}(c(t_q, \alpha, \beta); \lambda, \boldsymbol{\theta}),$$

where $F_{NMV}(\cdot, \lambda, \boldsymbol{\theta}) = F_{NMV}(\cdot; 0, \lambda, 1, \boldsymbol{\theta})$ is the standardized cdf of the NMV distribution.

Proof. The proof is trivial and has been omitted. □

Note that there are no closed expressions for the PS and PO of the NMV-GBS random variable. Therefore, the risks can be evaluated by using the function `pghyp( )` in the R package **ghyp**.

The TS risk measure is defined as the first order upper partial moment with respect to the threshold $t_q \in \mathbb{R}^+$. The next theorem provides a means of calculating the TS when T follows a NMV-GBS distribution.

Theorem 3. *Let $T \sim \mathcal{NMV}-\mathcal{GBS}(\alpha,\beta,\lambda,\boldsymbol{\theta})$. The TS of random variable T takes the form of*

$$\begin{aligned} TS_T(t_q,\alpha,\beta,\lambda,\boldsymbol{\theta}) &= (t_q-\beta)PS(t_q,\alpha,\beta,\lambda,\boldsymbol{\theta}) - \frac{\beta}{2}\omega_1 - E[(W+W^{5/2}\lambda^2)\Phi(c(t_q,\alpha,\beta);W\lambda,W)] \\ &\quad -E[(W\lambda^2 - c(t_q,\alpha,\beta) - 2W^2\lambda)\phi(c(t_q,\alpha,\beta);W\lambda,W)], \end{aligned}$$

where the last expectation is computed by (12) *and*

$$\omega_1 = \int_{c(t_q,\alpha,\beta)}^{\infty} \left(x\sqrt{\alpha^2x^2+4}\right) f_{NMV}(x;\lambda,\boldsymbol{\theta})\, dx.$$

Proof. The TS of random variable T is defined as

$$TS(t_q,\alpha,\beta,\lambda,\boldsymbol{\theta}) = \mathbb{E}[T-t_q]_+^1 = \int_{t_q}^{\infty}(t-t_q)f_{\text{NMV-GBS}}(t;\alpha,\beta,\lambda,\boldsymbol{\theta})\, dt = \int_{t_q}^{\infty} t f_{\text{NMV-GBS}}(t;\alpha,\beta,\lambda,\boldsymbol{\theta})\, dt - t_q PO(t_q,\alpha,\beta,\lambda,\boldsymbol{\theta}).$$

Using (4), the above integral can be rearranged as

$$\begin{aligned} \int_{t_q}^{\infty} t f_{\text{NMV-GBS}}(t;\alpha,\beta,\lambda,\boldsymbol{\theta})\, dt &= \int_{c(t_q,\alpha,\beta)}^{\infty} \frac{\beta}{4}\left[\alpha x + \sqrt{\alpha^2x^2+4}\right]^2 f_{\text{NMV}}(x;\lambda,\boldsymbol{\theta})\, dx \\ &= \beta PO(t_q,\alpha,\beta,\lambda,\boldsymbol{\theta}) + \frac{\beta}{2}\omega_1 + \frac{\alpha^2\beta}{2}\int_{c(t_q,\alpha,\beta)}^{\infty} x^2 f_{\text{NMV}}(x;\lambda,\boldsymbol{\theta})\, dx. \end{aligned}$$

Using representation (1), the last integral can be rewritten as

$$\begin{aligned} \int_{c(t_q,\alpha,\beta)}^{\infty} x^2 f_{\text{NMV}}(x;\lambda,\boldsymbol{\theta})\, dx &= \int_0^{\infty}\int_{c(t_q,\alpha,\beta)}^{\infty} x^2\phi(x;w\lambda,w)h(w;\boldsymbol{\theta})\, dx\, dw \\ &= \int_0^{\infty}\int_{c(t_q,\alpha,\beta)}^{\infty} \left(w\left(\frac{x-w\lambda}{\sqrt{w}}\right)^2 + 2w^{3/2}\lambda\left(\frac{x-w\lambda}{\sqrt{w}}\right) + w^2\lambda^2\right)\phi\left(\frac{x-w\lambda}{\sqrt{w}}\right)h(w;\boldsymbol{\theta})\, dx\, dw \\ &= \int_0^{\infty}\left(f_1(t_q)+f_2(t_q)+f_3(t_q)\right)h(w;\boldsymbol{\theta})\, dw, \end{aligned}$$

where

$$f_1(t_q) = \int_{c(t_q,\alpha,\beta)}^{\infty} w\left(\frac{x-w\lambda}{\sqrt{w}}\right)^2 \phi\left(\frac{x-w\lambda}{\sqrt{w}}\right) dx = w\Big((1-\Phi\left(c(t_q,\alpha,\beta);w\lambda,w\right)) + (c(t_q,\alpha,\beta)-w\lambda)\phi\left(c(t_q,\alpha,\beta);w\lambda,w\right)\Big),$$

$$f_2(t_q) = 2w^{3/2}\lambda\int_{c(t_q,\alpha,\beta)}^{\infty} w\left(\frac{x-w\lambda}{\sqrt{w}}\right)\phi\left(\frac{x-w\lambda}{\sqrt{w}}\right) dx = 2w^2\lambda\phi\left(c(t_q,\alpha,\beta);w\lambda,w\right),$$

and

$$f_3(t_q) = \int_{c(t_q,\alpha,\beta)}^{\infty} w^2\lambda^2\, \phi\left(\frac{x-w\lambda}{\sqrt{w}}\right) dx = w^{5/2}\lambda^2\left(1-\Phi\left(c(t_q,\alpha,\beta);w\lambda,w\right)\right).$$

Therefore, we have

$$\begin{aligned} \int_{t_q}^{\infty} t f_{\text{NMV-GBS}}(t;\alpha,\beta,\lambda,\boldsymbol{\theta})\, dt &= \beta PO(t_q,\alpha,\beta,\lambda,\boldsymbol{\theta}) + \frac{\beta}{2}\omega_1 + E_W[(W+W^{5/2}\lambda^2)(1-\Phi(c(t_q,\alpha,\beta);W\lambda,W))] \\ &\quad +E[(Wc(t_q,\alpha,\beta) - W^2\lambda^2 + 2W^2\lambda)\phi(c(t_q,\alpha,\beta);W\lambda,W)]. \end{aligned}$$

□

The VaR is a widely employed measure of downside risk in capital markets. Given a confidence level $q \in (0, 1)$, the VaR is defined as

$$\text{VaR}_q(T) = \sup\{t \mid F_T(t) \leq q\}. \tag{13}$$

However, VaR is often criticized for its lack of coherence properties because it is sensitive to the shape of the tail of the loss distribution. As an alternative, TVaR is a coherent risk measure that fulfills the properties of monotonicity, sub-additivity, homogeneity, and translational invariance and can be viewed as the expected worst. More precisely, TVaR gives the expected amount of extreme loss under a given risk. Given a confidence level $q \in (0, 1)$, the TVaR is defined as $\text{TVaR}(T; q) = E[T|T \geq \text{VaR}_q(T)]$, where $\text{VaR}_q(T)$ is the possible loss obtained by the qth percentile of T as defined in (13). We characterize the well-known tail conditional expectation, namely the TVaR measure, in Theorem 4.

Theorem 4. *Let $VaR_q(T) = t_q$. The TVaR measure of the random variable T distributed as $\mathcal{NMV}-\mathcal{GBS}(\alpha,\beta,\lambda,\boldsymbol{\theta})$ is given by*

$$\begin{aligned} TVaR(T;q) &= \frac{1}{1-q}\Big(\beta PO(t_q,\alpha,\beta,\lambda,\boldsymbol{\theta}) + \frac{\beta}{2}\omega_1 + E[(W + W^{5/2}\lambda^2)(1 - \Phi(c(t_q,\alpha,\beta); W\lambda, W))] \\ &\quad + E\big[(Wc(t_q,\alpha,\beta) - W^2\lambda^2 + 2W^2\lambda)\phi(c(t_q,\alpha,\beta); W\lambda, W)\big]\Big). \end{aligned}$$

Proof. By definition of the TVaR, we have

$$\text{TVaR}(T;q) = \frac{1}{1-q}\int_{t_q}^{\infty} t f_{\text{NMV-GBS}}(t;\alpha,\beta,\lambda,\boldsymbol{\theta})\, dt,$$

which completes the proof. □

Corollary 1. *Let T be distributed by a mixture of NMV-GBS distributions with pdf (8) and $VaR_q(T) = t_q$. Then, the TVaR measure of T is*

$$\begin{aligned} TVaR(T;q) &= \frac{1}{1-q}\sum_{i=1}^{g}\Bigg(\beta_i PO(t_q,\alpha_i,\beta_i,\lambda_i,\boldsymbol{\theta}_i) + \frac{\beta_i}{2}\omega_{i1} + E[(W_i + W_i^{5/2}\lambda_i^2)(1 - \Phi(c(t_q,\alpha_i,\beta_i); W_i\lambda, W_i))] \\ &\quad + E[(W_i c(t_q,\alpha_i,\beta_i) - W_i^2\lambda_i^2 + 2W_i^2\lambda_i)\phi(c(t_q,\alpha_i,\beta); W_i\lambda_i, W_i)]\Bigg). \end{aligned}$$

5. Application

5.1. Data description

In this section, we consider four economic real datasets including film revenues (Film rev.), the Munich rent (rent99) and the FTSE 100 Index (log-FTSE), which are all available in the R package "**gamlss.data**", and the exchange rate between US dollars (USD) and British pound sterling (Ex. US. UK), which is available in the R package "**CASdatasets**". The Film rev. data, in USD, was derived from standard industry data sourced by Nielsen EDI for the North American market for the period 1988 to 1999. The rent99 dataset, collected in the year 1999, contains the monthly rental price, known as the nett rent, which remains after having subtracted all running costs and incidentals-per square meter. The third dataset contains the natural logarithm of the daily returns from the international stock market, the FTSE 100. The Ex. US. UK are the daily buying rates in New York City for cable transfers payable in foreign currencies between January 4, 1971 and in March 1, 2013. Researchers had analyzed the datasets [61, 62] concluded that they cover different features such as strong skewness and leptokurtosis, right heavy-tail and bi-, as well as, multimodality. These characteristics motivate us to apply our proposed methodology to the data for illustrative purposes.

Table 2 provides summary statistics of the data, including the number of observations (n), mean, standard deviation (St.Dev), minimum (min), maximum (max), skewness (γ_x), kurtosis (κ_x) and the Jarque-Bera test statistic [63] along with its corresponding P-value. The results of the standard deviation reveal that the log-FTSE data is less risky than the others. The Jarque-Bera statistic and its extremely low P-value demonstrate significantly that the datasets do note meet normality assumptions. This issue can also be concluded from the normal quantile-quantile (Q-Q) plots given in Figure 2, depicting a point inflection with different slopes to the left and right.

Table 2: Descriptive statistics of the rent99, Film rev., Ex. US.UK and log-FTSE datasets.

	Measures								
Data	n	mean	St.Dev	min	max	γ_x	κ_x	Jarque-Bera	P-value
rent99	3082	7.111	2.436	0.416	17.722	0.299	-0.127	47.830	<4.109e-11
Film rev.	4031	11.783	3.068	4.212	18.068	0.037	-1.329	297.310	<2.2e-16
Ex. US. UK	10583	1.772	0.313	1.052	2.644	0.911	0.136	1473.000	<2.2e-16
log-FTSE	1000	8.760	0.067	8.568	8.868	-0.730	-0.364	94.685	< 2.2e-16

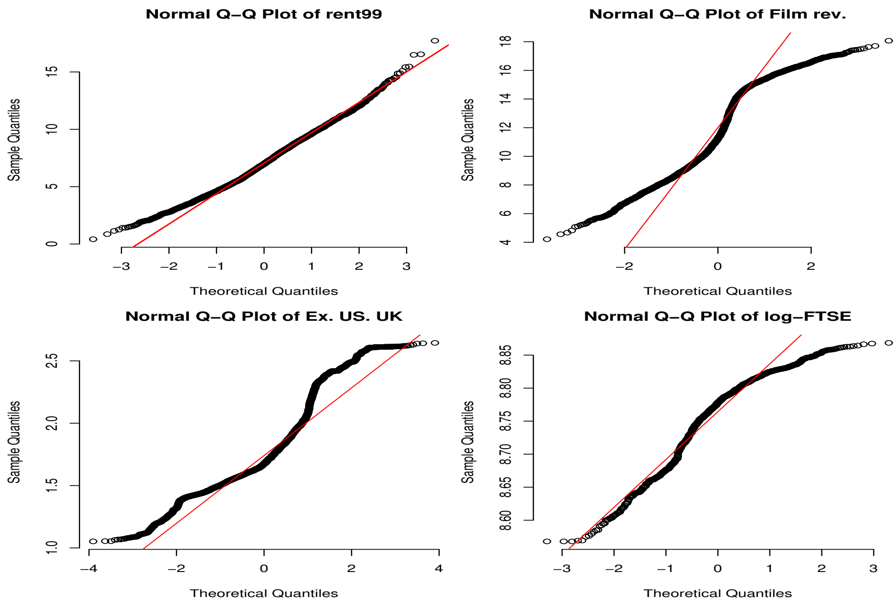

Figure 2: Normal Q-Q plots for the rent99, Film rev., Ex. US.UK and log-FTSE datasets.

5.2. *Model evaluation*

The results of the previous section motivate us to consider skew distributions for more accurate analysis. We fit special cases of the FM-NMV-GBS distributions described in Section 3 and summarized in Table 1 as alternative benchmark models to these economic datasets. The ECME algorithm for estimating the model parameters is implemented by exploiting computational aspects of initial points, convergence rule, and model selection criterion. We fit all models for g ranging from 1 to 6 and find the best choice of the mixing components number g based on the BIC criterion for each dataset.

Table 3 displays the log-likelihood, along with the BIC values (for the best g) obtained by fitting the nine considered models to each dataset. The results based on the BIC indicate that the finite mixture of strongly skewed and right heavily-tailed NMV-GBS distributions provide a highly improved fit for the data. The rank of models based on the BIC from 1 to 9 is presented in Table 3. It can be seen that the FM-H-BS, FM-VG-BS, FM-NIG-BS, FM-GHST-BS and FM-NMVBS-BS models are almost ranked from 1 to 5.

To get reliable risk measures, it is crucial to verify the validity of a model in terms of goodness-of-fit tests. We perform the Kolmogorov-Smirnov's (KS) goodness-of-fit test to check the similarity assessment of the experimental data against the fitted distributions. Table 3 depicts the resulting KS test. It can be observed that the P-values of the KS test for the FM-H-BS, FM-VG-BS, FM-NIG-BS, FM-GHST-BS, FM-NMVBS-BS and FM-NMVL-BS models are greater than the 5% significance level, reflecting that these data strongly follow a FM-NMV-GBS distributions. Figure 3 shows the histograms of data overlaid with the best fitted curves together with their probability-probability (P-P) plots. The bimodality of the data and the suitability of the best model to fit the data can be observed.

Table 3: Model comparison criteria of the fitted models to the considered datasets.

Data	g	Criteria	L-BS	T-BS	SL-BS	H-BS	VG-BS	NIG-BS	GHST-BS	NMVBS-BS	NMVL-BS
		m	5	7	7	9	11	9	8	9	9
rent99	2	$\ell(\hat{\Theta})$	-7135.65	-7074.74	-7093.19	-7066.10	-7064.99	-7064.73	-7064.28	-7065.13	-7082.89
		BIC	14311.46	14205.71	14242.61	14204.49	14218.35	14201.76	14192.82	14202.56	14238.08
		KS	0.0244	0.0113	0.0203	0.0056	0.0063	0.0068	0.0072	0.0070	0.0156
		P-value	0.1238	0.8146	0.2728	0.9999	0.9994	0.9985	0.9975	0.9982	0.4382
		Rank	9	5	8	4	6	2	1	3	7
		m	5	7	7	9	11	9	8	9	9
Film rev.	2	$\ell(\hat{\Theta})$	-9518.18	-9398.06	-9473.57	-9383.78	-9385.68	-9383.34	-9382.43	-9397.21	-9453.00
		BIC	19077.87	18854.24	19005.25	18842.28	18862.68	18841.40	18831.27	18869.13	18980.72
		KS	0.0249	0.0084	0.0214	0.0087	0.0091	0.0085	0.0083	0.0096	0.0198
		P-value	0.0712	0.9336	0.0814	0.9228	0.8790	0.9282	0.9278	0.8564	0.1823
		Rank	9	4	8	3	5	2	1	6	7
		m	11	15	15	19	23	19	16	19	19
Ex. US. UK	4	$\ell(\hat{\Theta})$	-537.38	-463.35	-388.27	-342.90	-325.75	-369.12	-382.23	-366.53	-371.64
		BIC	1176.69	1065.71	915.55	861.87	864.66	914.32	912.72	909.14	919.35
		KS	0.0172	0.0139	0.0153	0.0087	0.0112	0.0091	0.0102	0.0084	0.0122
		P-value	0.0042	0.0322	0.0161	0.3926	0.1934	0.3328	0.2225	0.4366	0.1174
		Rank	9	8	6	1	2	5	4	3	7
		m	8*	7	7	9	11	9	8	9	9
log-FTSE	2	$\ell(\hat{\Theta})$	1414.56	1418.47	1423.42	1436.48	1436.62	1435.53	1429.51	1436.32	1430.90
		BIC	-2773.86	-2788.58	-2798.49	-2810.80	-2797.25	-2808.88	-2803.75	-2810.47	-2799.63
		KS	0.0234	0.0280	0.0431	0.0193	0.0153	0.0200	0.0189	0.0206	0.0278
		P-value	0.6226	0.4132	0.1523	0.8640	0.9746	0.8146	0.8965	0.8058	0.4434
		Rank	9	8	6	1	7	3	4	2	5

*: In this case the best number of mixture component is $g = 3$.

5.3. Application to some risk measures

We utilize the parameter estimates obtained from the previous section to compare the accuracy of predicted VaR and TVaR values. We generate two millions samples from each model to evaluate VaR and TVaR. Recall that the VaR is the qth percentile of the simulated loss samples, whereas TVaR is the mean loss and is thus greater than VaR. Table 4 presents the detailed numerical results of the estimated VaR and TVaR based on various models fitted to the four considered datasets at confidence levels of 99%, 97.5%, and 95%. It can be observed that the skewed and heavy-tailed sub-models of the FM-NMV-GBS distributions provide closer prediction of the true VaR and TVaR values in most cases. In order to provide visual comparison, Figure 4 displays the empirical values overlaid with theoretical predicted VaR and TVaR of the five best fitted models for confidence levels ranging between 90% and 99%. The lines in these figures highlight that the predicted VaRs and TVaRs are very close to the empirical ones.

To assess relative changes on the theoretical predictions, we also calculate the mean absolute relative error (MARE) defined as

$$\text{MARE} = \frac{1}{n_q}\sum_{l=1}^{n_q}\left|\frac{M_l - \hat{M}_l}{M_l}\right|,$$

where M_l and $\hat{M}_l$ are the lth empirical and its corresponding theoretical predicted risk measures, respectively, for the $n_q = 30$ confidence level chosen from the interval (0.9,1). Table 5 summarizes the results of MARE to assess relative changes on the theoretical prediction. The results depicted in Table 5 reveal that the FM-H-BS, FM-VG-BS, FM-NIG-BS, FM-GHST-BS and FM-NMVBS-BS models mostly show a lesser amount of MARE, thus estimating the VaR and TVaR more accurately.

6. Conclusion and future extensions

Non-normal features, such as asymmetry and heavy tails, are often present in economic and financial data streams. To analyze these data, the family of NMV distributions can be a good candidate for risk management. The outline

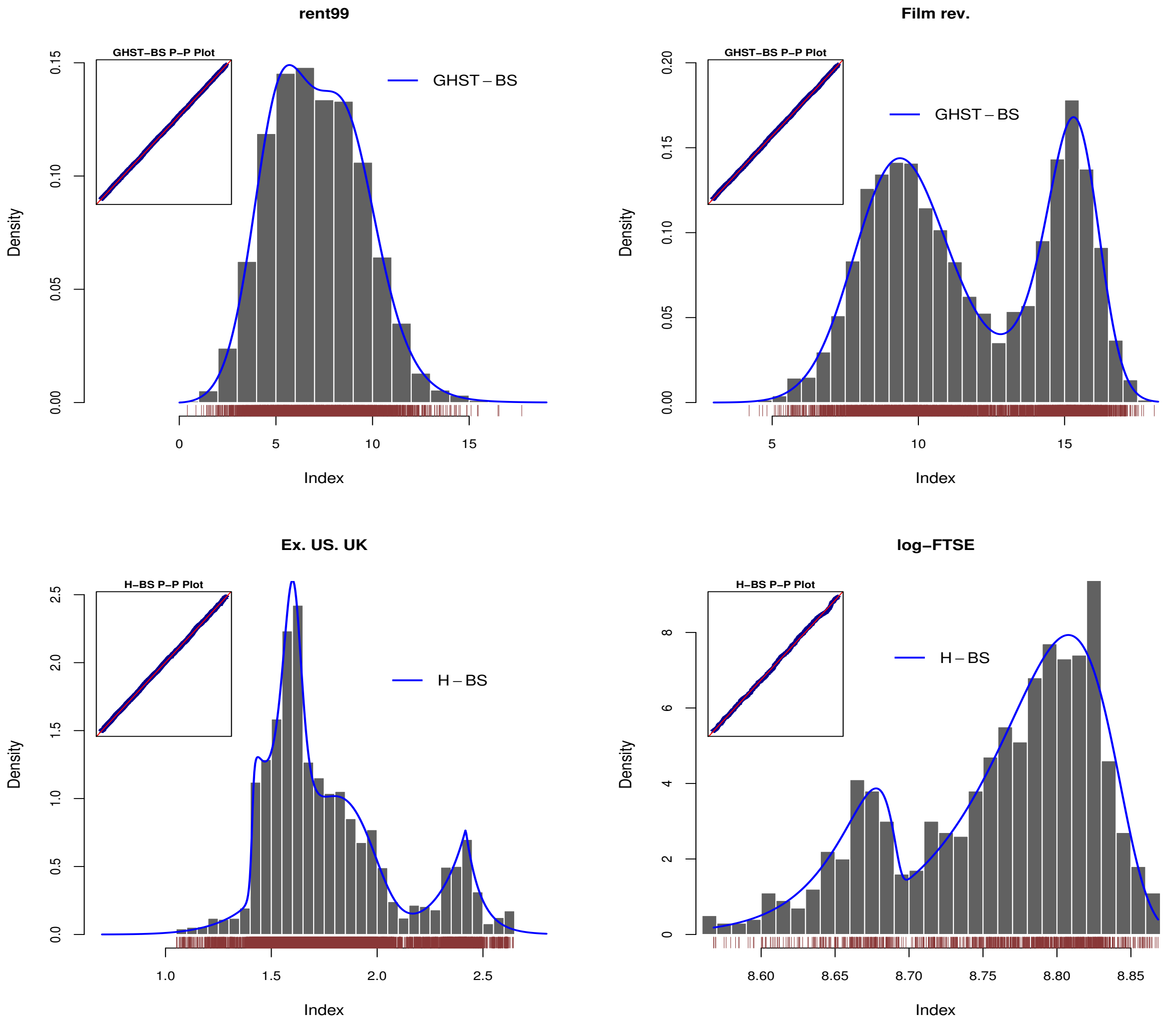

Figure 3: Histogram overlaid with the best fitted density and P-P plots for the rent99, Film rev., Ex. US.UK, and log-FTSE datasets.

of this paper has been divided into three parts. The first one has dealt with proposing a new generalization of the BS distribution based on the NMV model and its well-known sub-class, the GH family. In the second part, we have introduced a promising finite mixture model based on the new extension of BS distribution (8) for analyzing and clustering positive valued data. Some mathematical, statistical and financial properties of the new model have also been derived. The feasible ECME algorithm has been developed for calibrating the parameters of the proposed finite mixture model to the data. Finally, in the last part, the application to real datasets has been presented. The numerical results obtained by fitting the proposed class of finite mixture models to four real examples suggest that five special cases of the FM-GH-GBS distribution, especially the FM-GHST-BS and FM-H-BS models, outperform the others. Moreover, it is shown that the FM-NMV-GBS distributions is well suited for risk measurements.

The methodology presented in this paper could facilitate the development of model-based clustering for the multivariate [44] as well as matrix variate [64] positive data. Another possible extension of this work is to consider the developing a time series model similar to that of [65], and the ARMA and GARCH ones as more interesting cases, based on the NMV-GBS distribution.

All computations were carried out using R 3.4.3 in a Win 64 environment with a 2.50 GHz/Intel Core(TM) i3 3120M CPU Processor and 4.0 GB RAM. The R code is available from the second author upon request.

Table 4: Comparison of estimated VaR and TVaR for four real economic datasets based on the FM-NMV-GBS sub-models at 99%, 975%, and 95% confidence levels.

Data	Criteria	Level	Empir.	L-BS	T-BS	SL-BS	H-BS	VG-BS	NIG-BS	GHST-BS	NMVBS-BS	NMVL-BS
rent99	VaR	0.99	13.077	15.653	13.586	13.885	13.068	13.059	13.070	13.072	13.070	13.626
		0.975	11.947	13.223	12.233	12.250	11.991	12.005	12.023	12.007	12.023	12.147
		0.95	11.222	11.675	11.292	11.150	11.170	11.190	11.198	11.187	11.198	11.123
	TVaR	0.99	14.305	19.285	17.199	16.100	14.325	14.231	14.240	14.274	14.242	15.540
		0.975	13.167	16.243	14.553	14.196	13.198	13.165	13.178	13.182	13.178	13.874
		0.95	12.339	14.293	13.130	12.912	12.367	12.360	12.372	12.367	12.373	12.722
Film rev.	VaR	0.99	16.920	18.182	17.061	17.114	16.887	16.897	16.883	16.889	17.038	17.106
		0.975	16.523	17.071	16.603	16.547	16.520	16.522	16.521	16.525	16.594	16.549
		0.95	16.210	16.376	16.220	16.168	16.197	16.198	16.198	16.200	16.216	16.171
	TVaR	0.99	17.172	20.244	17.688	18.272	17.273	17.334	17.246	17.243	17.556	18.211
		0.975	16.902	18.609	17.155	17.377	16.917	16.944	16.906	16.904	17.094	17.352
		0.95	16.622	17.643	16.774	16.856	16.631	16.645	16.626	16.627	16.741	16.844
Ex. US. UK	VaR	0.99	2.583	2.573	2.553	2.550	2.549	2.565	2.551	2.496	2.550	2.550
		0.975	2.485	2.489	2.492	2.479	2.479	2.474	2.486	2.456	2.483	2.480
		0.95	2.419	2.432	2.433	2.428	2.427	2.424	2.429	2.416	2.428	2.428
	TVaR	0.99	2.609	2.696	2.604	2.635	2.631	2.684	2.620	2.532	2.630	2.631
		0.975	2.560	2.593	2.553	2.559	2.557	2.580	2.556	2.497	2.557	2.558
		0.95	2.504	2.524	2.507	2.505	2.503	2.512	2.506	2.466	2.505	2.505
log-FTSE	VaR	0.99	8.860	8.877	8.867	8.861	8.859	8.861	8.859	8.859	8.859	8.861
		0.975	8.852	8.858	8.854	8.849	8.850	8.850	8.851	8.851	8.851	8.849
		0.95	8.843	8.845	8.844	8.840	8.843	8.841	8.843	8.843	8.843	8.840
	TVaR	0.99	8.864	8.899	8.878	8.875	8.866	8.872	8.866	8.866	8.865	8.874
		0.975	8.860	8.879	8.867	8.863	8.859	8.861	8.859	8.859	8.859	8.862
		0.95	8.854	8.865	8.858	8.853	8.852	8.853	8.853	8.853	8.852	8.853

Table 5: Comparison of estimation accuracy of VaR and TVaR in terms of MARE (%).

	rent99		Film rev.		Ex. US. UK		log-FTSE	
FM-model	VaR	TVaR	VaR	TVaR	VaR	TVaR	VaR	TVaR
L-BS	5.5906	16.8302	2.1474	7.8439	0.4187	1.0137	0.0464	0.1530
T-BS	1.2388	7.3574	0.2895	1.2160	0.4040	0.1481	0.0282	0.0613
SL-BS	1.7660	5.0243	0.3134	2.2295	0.3091	0.1028	0.0249	0.0181
H-BS	0.2892	0.1133	0.0793	0.1351	0.2981	0.0794	0.0235	0.0082
VG-BS	0.2550	0.1067	0.0776	0.2638	0.3267	0.0965	0.0204	0.0124
NIG-BS	0.2864	0.1644	0.0777	0.0813	0.3255	0.0840	0.0248	0.0080
GHST-BS	0.2692	0.1079	0.0773	0.0855	0.3307	0.0988	0.0237	0.0075
NMVBS-BS	0.2885	0.1662	0.2599	0.9208	0.3119	0.0706	0.0256	0.0083
NMVL-BS	1.4066	3.3353	0.2983	2.1099	0.2959	0.0898	0.0201	0.0173

Acknowledgements

The authors wish to thank the Editor-in-Chief, anonymous Associate Editor, and referees for their helpful comments, which helped to improve this article. This work was partially supported by the National Research Foundation, South Africa (Reference: CPRR160403161466, Grant No. 105840, SARChI Research Chair- UID: 71199, and STATOMET).

References

[1] T. Bodnar, A. K. Gupta, Robustness of the inference procedures for the global minimum variance portfolio weights in a skew-normal model, The European Journal of Finance 21 (13-14) (2015) 1176–1194.

[2] R. Vernic, Multivariate skew-normal distributions with applications in insurance, Insurance: Mathematics and Economics 38 (2) (2006) 413–426.

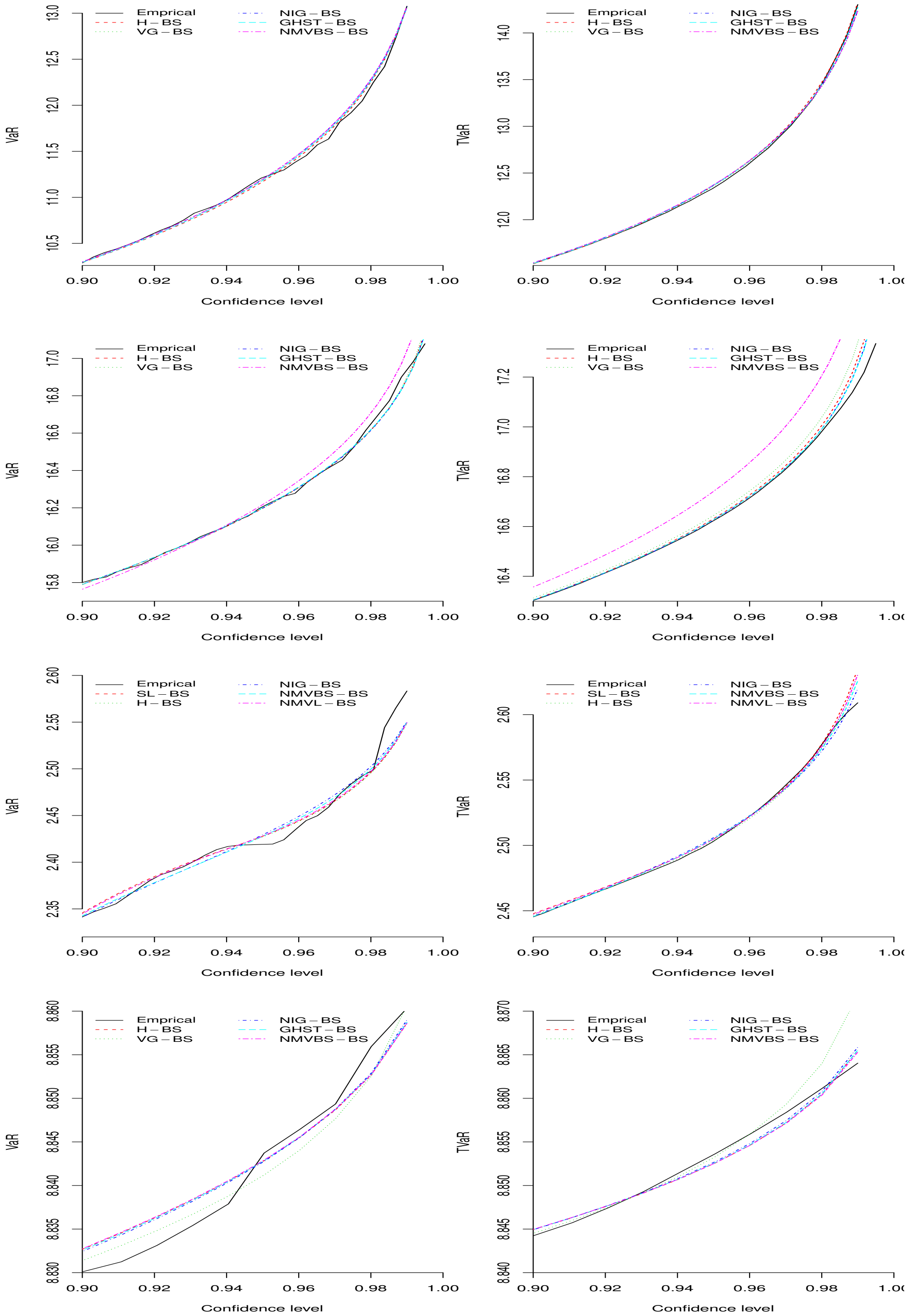

Figure 4: VaR (left panel) and TVaR (right panel) plots as a function of confidence levels (q) for the rent99 (1st), Film rev. (2nd), Ex. US.UK (3th) and log-FTSE (4th) datasets from top to bottom.

[3] A. Azzalini, A class of distributions which includes the normal ones, Scandinavian Journal of Statistics 12 (2) (1985) 171–178.

[4] M. Eling, Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models?, Insurance: Mathematics and Economics 51 (2) (2012) 239–248.
[5] M. Eling, Fitting asset returns to skewed distributions: Are the skew-normal and skew-student good models?, Insurance: Mathematics and Economics 59 (2014) 45–56.
[6] A. Azzalini, A. Capitanio, Statistical applications of the multivariate skew normal distribution, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61 (3) (1999) 579–602.
[7] T. Shushi, Skew-elliptical distributions with applications in risk theory, European Actuarial Journal 7 (1) (2017) 277–296.
[8] K. Prause, et al., The generalized hyperbolic model: Estimation, financial derivatives and risk measures, Ph.D. thesis, Verlag nicht ermittelbar (1999).
[9] K. Aas, The generalized hyperbolic skew Student's *t* distribution, Journal of Financial Econometrics 4 (2) (2006) 275–309.
[10] M. Bee, M. M. Dickson, F. Santi, Likelihood-based risk estimation for Variance-Gamma models, Statistical Methods & Applications 27 (1) (2017) 69–89.
[11] A. Azzalini, T. d. Cappello, S. Kotz, Log-skew-normal and log-skew-*t* distributions as models for family income data, Journal of Income Distribution 11 (3) (2003) 12.
[12] A. B. Salem, T. D. Mount, A convenient descriptive model of income distribution: the Gamma density, Econometrica: Journal of the Econometric Society (1974) 1115–1127.
[13] S. X. Chen, Probability density function estimation using Gamma kernels, Annals of the Institute of Statistical Mathematics 52 (3) (2000) 471–480.
[14] A. Punzo, L. Bagnato, A. Maruotti, Compound unimodal distributions for insurance losses, Insurance: Mathematics and Economics 81 (2018) 95–107.
[15] A. Mazza, A. Punzo, Modeling household income with contaminated unimodal distributions, in: New Statistical Developments in Data Science, Springer, 2019, pp. 373–391.
[16] L.-G. Benckert, The Lognormal model for the distribution of one claim, ASTIN Bulletin 2 (01) (1962) 9–23.
[17] A. Punzo, A new look at the inverse Gaussian distribution with applications to insurance and economic data, Journal of Applied Statistics 46 (7) (2018) 1260–1287.
[18] O. L. Gebizlioglu, B. Şenoğlu, Y. M. Kantar, Comparison of certain Value-at-Risk estimation methods for the two-parameter Weibull loss distribution, Journal of Computational and Applied Mathematics 235 (11) (2011) 3304–3314.
[19] Z. Birnbaum, S. Saunders, A new family of life distributions, Journal of Applied Probability 6 (02) (1969) 319–327.
[20] V. Leiva, M. Barros, G. A. Paula, M. Galea, Influence diagnostics in Log-Birnbaum-Saunders regression models with censored data, Computational Statistics & Data Analysis 51 (12) (2007) 5694–5707.
[21] S. A. Bakar, N. Hamzah, M. Maghsoudi, S. Nadarajah, Modeling loss data using composite models, Insurance: Mathematics and Economics 61 (2015) 146–154.
[22] K. Mohammadi, O. Alavi, J. G. McGowan, Use of Birnbaum-Saunderss distribution for estimating wind speed and wind power probability distributions: A review, Energy Conversion and Management 143 (2017) 109–122.
[23] C. R. Bhatti, The Birnbaum-Saunders autoregressive conditional duration model, Mathematics and Computers in Simulation 80 (10) (2010) 2062–2078.
[24] X. Jin, J. Kawczak, et al., Birnbaum-Saunders and Log-normal kernel estimators for modelling durations in high frequency financial data, Annals of Economics and Finance 4 (2003) 103–124.
[25] G. A. Paula, V. Leiva, M. Barros, S. Liu, Robust statistical modeling using the Birnbaum-Saunders-*t* distribution applied to insurance, Applied Stochastic Models in Business and Industry 28 (1) (2011) 16–34.
[26] F. Hashemi, M. Naderi, M. Mashinchi, Clustering right-skewed data stream via Birnbaum–Saunders mixture models: A flexible approach based on fuzzy clustering algorithm, Applied Soft Computing 82 (2019) 105539.
[27] D. Kundu, N. Kannan, N. Balakrishnan, On the hazard function of Birnbaum-Saunders distribution and associated inference, Computational Statistics & Data Analysis 52 (5) (2008) 2692–2702.
[28] G. M. Cordeiro, A. J. Lemonte, E. M. Ortega, An extended fatigue life distribution, Statistics 47 (3) (2013) 626–653.
[29] G. M. Cordeiro, A. J. Lemonte, The exponentiated generalized Birnbaum-Saunders distribution, Applied Mathematics and Computation 247 (2014) 762–779.
[30] A. J. Lemonte, A new extended Birnbaum-Saunders regression model for lifetime modeling, Computational Statistics & Data Analysis 64 (2013) 34–50.
[31] M. Bourguignon, R. B. Silva, G. M. Cordeiro, A new class of fatigue life distributions, Journal of Statistical Computation and Simulation 84 (12) (2014) 2619–2635.
[32] M. Khosravi, V. Leiva, A. Jamalizadeh, E. Porcu, On a nonlinear Birnbaum-Saunders model based on a bivariate construction and its characteristics, Communications in Statistics - Theory and Methods 45 (3) (2015) 772–793.
[33] R. G. Romeiro, F. Vilca, N. Balakrishnan, A robust multivariate Birnbaum-Saunders distribution: EM estimation, Statistics 52 (2) (2017) 321–344.
[34] F. Hashemi, M. Naderi, A. Jamalizadeh, Normal mean-variance Lindley Birnbaum-Saunders distribution, Statistics and Its Interface 12 (2019) 585–597.
[35] K. Burnecki, J. Janczura, R. Weron, Building loss models, in: Statistical Tools for Finance and Insurance, Springer Berlin Heidelberg, 2011, pp. 293–328.
[36] A. Punzo, Discrete beta-type models, in: Locarek-Junge H., Weihs C. (Eds.), Classification as a Tool for Research, Studies in Classification, Data Analysis, and Knowledge Organization, Berlin Heidelberg: Springer-Verlag, 2010, pp. 253–261.
[37] T. Miljkovic, B. Grün, Modeling loss data using mixtures of distributions, Insurance: Mathematics and Economics 70 (2016) 387–396.
[38] L. Bagnato, A. Punzo, Finite mixtures of unimodal Beta and Gamma densities and the *k*-bumps algorithm, Computational Statistics 28 (4) (2012) 1571–1597.
[39] M. Naderi, M. Mozafari, K. Okhli, Finite mixture modeling via skew-Laplace Birnbaum-Saunders distribution, Journal of Statistical Theory

and Applications (Accepted).

[40] S. D. Tomarchio, A. Punzo, Modelling the loss given default distribution via a family of zero-and-one inflated mixture models, Journal of the Royal Statistical Society: Series A (Statistics in Society) 182 (4) (2019) 1247–1266.

[41] A. J. McNeil, R. Frey, P. Embrechts, et al., Quantitative risk management: Concepts, techniques and tools, Vol. 3, Princeton university press Princeton, 2005.

[42] H. Markowitz, Portfolio selection, The Journal of Finance 7 (1) (1952) 77–91.

[43] B. Jørgensen, Statistical properties of the generalized inverse Gaussian distribution, Vol. 9, Springer Science & Business Media, 1982.

[44] N. Balakrishnan, D. Kundu, Birnbaum-Saunders distribution: A review of models, analysis, and applications, Applied Stochastic Models in Business and Industry 35 (1) (2018) 4–49.

[45] A. Jamalizadeh, F. Hashemi, M. Naderi, Discussion of "Birnbaum-Saunders distribution: A review of models, analysis, and applications", Applied Stochastic Models in Business and Industry 35 (1) (2018) 82–89.

[46] A. J. Lemonte, G. M. Cordeiro, Discussion of "Birnbaum-Saunders distribution: A review of models, analysis, and applications", Applied Stochastic Models in Business and Industry 35 (1) (2018) 56–60.

[47] V. Leiva, F. Vilca, N. Balakrishnan, A. Sanhueza, A skewed sinh-normal distribution and its properties and application to air pollution, Communications in Statistics - Theory and Methods 39 (3) (2010) 426–443.

[48] M. Naderi, A. Arabpour, T.-I. Lin, A. Jamalizadeh, Nonlinear regression models based on the normal mean-variance mixture of Birnbaum-Saunders distribution, Journal of the Korean Statistical Society 46 (3) (2017) 476–485.

[49] M. Naderi, W. L. Hung, T. I. Lin, A. Jamalizadeh, A novel mixture model using the multivariate normal mean-variance mixture of Birnbaum-Saunders distributions and its application to extrasolar planets, Journal of Multivariate Analysis 171 (2019) 126–138.

[50] M. Naderi, A. Arabpour, A. Jamalizadeh, Multivariate normal mean-variance mixture distribution based on Lindley distribution, Communications in Statistics - Simulation and Computation 47 (4) (2017) 1179–1192.

[51] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society. Series B (Methodological) 39 (1) (1977) 1–38.

[52] C. Liu, D. B. Rubin, The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence, Biometrika 81 (4) (1994) 633–648.

[53] X.-L. Meng, D. B. Rubin, Maximum likelihood estimation via the ECM algorithm: A general framework, Biometrika 80 (2) (1993) 267–278.

[54] H. Ng, D. Kundu, N. Balakrishnan, Modified moment estimation for the two-parameter Birnbaum-Saunders distribution, Computational Statistics & Data Analysis 43 (3) (2003) 283–298.

[55] A. C. Aitken, On Bernoulli's numerical solution of algebraic equations, Proceedings of the Royal Society of Edinburgh 46 (1927) 289–305.

[56] P. D. McNicholas, T. B. Murphy, A. F. McDaid, D. Frost, Serial and parallel implementations of model-based clustering via parsimonious Gaussian mixture models, Computational Statistics & Data Analysis 54 (3) (2010) 711–723.

[57] D. Böhning, E. Dietz, R. Schaub, P. Schlattmann, B. G. Lindsay, The distribution of the likelihood ratio for mixtures of densities from the one-parameter exponential family, Annals of the Institute of Statistical Mathematics 46 (2) (1994) 373–388.

[58] B. G. Lindsay, Mixture models: theory, geometry, and applications, volume 5 of nsf-cbms regional conference series in probability and statistics, Institute for Mathematical Statistics: Hayward, CA.

[59] G. Schwarz, et al., Estimating the dimension of a model, The annals of statistics 6 (2) (1978) 461–464.

[60] A. J. Brogan, S. Stidham Jr, Non-separation in the mean-lower-partial-moment portfolio optimization problem, European Journal of Operational Research 184 (2) (2008) 701–710.

[61] V. Voudouris, R. Gilchrist, R. Rigby, J. Sedgwick, D. Stasinopoulos, Modelling skewness and kurtosis with the BCPE density in GAMLSS, Journal of Applied Statistics 39 (6) (2012) 1279–1293.

[62] A. Punzo, A. Mazza, A. Maruotti, Fitting insurance and economic data with outliers: a flexible approach based on finite mixtures of contaminated Gamma distributions, Journal of Applied Statistics 45 (14) (2018) 2563–2584.

[63] C. M. Jarque, A. K. Bera, A test for normality of observations and regression residuals, International Statistical Review/Revue Internationale de Statistique (1987) 163–172.

[64] L. Sánchez, V. Leiva, F. J. Caro-Lopera, F. J. A. Cysneiros, On matrix-variate Birnbaum-Saunders distributions and their estimation and application, Brazilian Journal of Probability and Statistics 29 (4) (2015) 790–812.

[65] T. Rahul, N. Balakrishnan, N. Balakrishna, Time series with Birnbaum-Saunders marginal distributions, Applied Stochastic Models in Business and Industry 34 (4) (2018) 562–581.