Statistics and Its Interface Volume 12 (2019) 585–597

Normal mean-variance Lindley Birnbaum-Saunders distribution

Farzane Hashemi, Mehrdad Naderi*, and Ahad Jamalizadeh

The generalization of Birnbaum-Saunders (BS) distribution has recently received considerable attention to provide accurate inferential results in dealing with survival data, reliability problems, fatigue life studies and hydrological data. This paper introduces a new extension of the BS distribution based on the normal mean-variance mixture of Lindley distribution. Since the proposed lifetime distribution can take positive and negative skewness and can have decreasing, increasing, upside-down bathtub, increasing-decreasing-increasing and decreasing-increasing-decreasing hazard rate functions, it may provide more flexible model than the existing extensions of BS distribution. Some properties of the new distribution are derived and the computationally analytical EM-type algorithm is developed for computing maximum likelihood estimates. Finally, the performance of the proposed methodology is illustrated through analyzing two real data sets.

AMS 2000 subject classifications: Primary 62E10, 65C60; secondary 62E15.
Keywords and phrases: Birnbaum-Saunders distribution, ECM-algorithm, Lindley distribution, Normal mean-variance mixture distribution.

1. INTRODUCTION

Among the lifetime models, the BS distribution [9] is perhaps the most widely considered and precious distribution in applied statistics. One of the main reasons for its importance is that the BS distribution has the advantage of considering the basic characteristics of the fatigue process. Moreover, the BS distribution is closely related to the normal distribution through a simple stochastic representation. Specifically, the random variable T is said to have a BS distribution with the shape and scale parameters α and β, respectively, if it can be generated by the following stochastic representation:

$$T = \frac{\beta}{4}\left[\alpha Y + \sqrt{(\alpha Y)^2 + 4}\right]^2, \tag{1}$$

where Y has a standard normal distribution. It can be easily seen that the probability density function (pdf) of T is given by

$$f(t;\alpha,\beta) = A(t,\alpha,\beta)\phi(a(t,\alpha,\beta)), \quad t,\ \alpha,\ \beta > 0, \tag{2}$$

where $\phi(\cdot)$ is the pdf of the standard normal distribution, $a(t,\alpha,\beta) = (\sqrt{t/\beta} - \sqrt{\beta/t})/\alpha$, and $A(t,\alpha,\beta)$ is the derivative of $a(t,\alpha,\beta)$ with respect to t. Although the BS distribution originally developed from a problem of material fatigue [9], it is recently applied in various areas of science and in diverse fields such as econometrics, biology, genetics, and engineering. For instance, some applications of the BS distribution were presented in [24, 31, 32, 6, 41].

The BS distribution is often criticized for its lack of robustness against atypical observations exhibiting highly skewness and heavy tail. Furthermore, the other major weakness of the BS distribution relies on the fact that its hazard rate function can only be upside-down bathtub-shaped [23] and it can not accommodate monotone (increasing or decreasing) or bathtub-shaped. This may cause misleading inferential results in practical applications since many real lifetime data exhibit monotone or bathtub-shaped hazard rate function. To overcome these deficiencies, several generalizations and extensions of the BS distribution have been recently introduced. For example, Corderio and Lemonte [14] proposed the exponentiated generalized BS distribution and studied its properties. Also, the five-parameter McDonald-Birnbaum-Saunders distribution were explored by Cordeiro et al. [13], which is an extension of the β-Birnbaum-Saunders distribution [12]. Some other recent works on extending BS distribution can be found in [11, 15, 19, 30, 42].

Another important approach of generalizing BS distribution is to replace the standard normal variable Y in (1) by other random variables followed by highly skewed and heavy tailed distributions, or by replacing $\phi(\cdot)$ in (2) by other pdf of asymmetric distributions. Based on this approach, Díaz-García and Leiva-Sánchez [17] introduced the generalized Birnbaum-Saunders distribution through assuming the general elliptical distributions. Considering the so-called skew-normal (SN) distribution [3], Vilca-Labra and Leiva-Sánchez [44] also proposed the skew-normal Birnbaum-Saunders (SN-BS) distribution. Vilca et al. [45] studied the SN-BS distribution properties and showed that the rang of skewness and kurtosis of the new model is higher than the BS distribution. Moreover, the skew-t Birnbaum-Saunders (ST-BS) and skew-normal-t Birnbaum-Saunders (SNT-BS) distributions were respectively introduced by Khosravi et al. [22] and Jamalizadeh et al. [18] via replacing skew-t and skew-normal-t distributions in (2).

*Corresponding author.

The class of normal mean-variance (NMV) mixture distribution [33] is an attractive alternate family of skewed distributions which contains both symmetric and asymmetric models as special cases. The class of NMV distributions have been recently considered as a good alternate for adequately modeling financial data. For example, Aas [1] demonstrated the superiority of two special cases of GH distribution, namely normal inverse Gaussian and generalized hyperbolic skew-t distributions, in evaluating Value-at-Risk (VaR) and Tail-Value-at-Risk (TVaR) measures. Bee et al. [7] also developed three extensions of the expectation-maximization (EM; [16]) type algorithm for obtaining maximum likelihood estimate of variance-Gamma distribution. They showed that the VG distribution is flexible enough to accommodate skewness and leptokurtosis in order to modeling log-returns of financial assets. Regarding possibly skewed and heavy-tailed data, Naderi et al. [38] proposed the normal mean-variance mixture of Lindley (NMVL) distribution and showed that the new model can be considered as a good alternative to existing skewed distributions for accurate inference. However, practical application of models with $\mathbb{R}$ support to the data defined on $\mathbb{R}^+$ may cause boundary bias, allocation of probability mass outside the theoretical support [43].

Considering the aforementioned properties of NMV and NMVL distributions and some practical limitations of BS model, the main objective of this paper is to introduce a new extension of the BS distribution via assuming the pdf of NMVL model in (2). We show that the proposed new model, called the normal mean-variance Lindley Birnbaum-Saunders distribution (NMVL-BS), has both monotonic and non-monotonic hazard rate function and has wide range of skewness and kurtosis. Also, we study some mathematical properties of the NMVL-BS distribution. To compute maximum likelihood (ML) estimate of model parameters, the Expectation Conditional Maximization (ECM) algorithm [36] is developed. Furthermore, we offer an information-based method for obtaining the asymptotic standard errors of the ML estimates. The validity of the proposed distribution in evaluating two risk measures, VaR and TVaR, is illustrated by analizing the fire loss data.

The rest of the paper is organised as follows. Section 2 provides a brief review of the NMVL distribution. In Section 3, the NMVL-BS distribution is introduced. Some mathematical properties and special cases of the new model are presented in Section 4. We also develop an ECM-type algorithm for ML estimation and provide a general information-based method for obtaining the asymptotic standard errors of ML estimates in Section 5. Section 6 gives two applications on real data to illustrate the performance of the proposed methodology. Two simulations are conducted, in Section 7, to examine the effect of outliers on model selection and to check the finite-sample property of ML estimates. We close the paper with a short summary in Section 8.

2. PRELIMINARY

A random variable X is followed by a generalized hyperbolic (GH) distribution if its pdf is given by

$$f_{\text{GH}}(x;\mu,\lambda,\sigma^2,\kappa,\chi,\psi) = C\exp\left\{(x-\mu)\lambda/\sigma^2\right\} \times K_{\kappa-0.5}\left(\sqrt{\left(\psi+\lambda^2/\sigma^2\right)\left(\chi+\delta(x,\mu,\sigma^2)\right)}\right) / \left(\sqrt{\left(\psi+\lambda^2/\sigma^2\right)\left(\chi+\delta(x,\mu,\sigma^2)\right)}\right)^{0.5-\kappa},$$

where $(x,\mu,\lambda,\kappa) \in R^4$, $K_\kappa(\cdot)$ denotes the third kind of the modified Bessel function with index κ, $\delta(x,\mu,\sigma^2) = \left(\frac{x-\mu}{\sigma}\right)^2$, the normalizing constant is

$$C = \frac{(\psi/\chi)^{\kappa/2}(\psi+(\lambda/\sigma)^2)^{0.5-\kappa}}{\sqrt{2\pi\sigma^2}K_\kappa(\sqrt{\psi\chi})},$$

and two parameters χ,ψ are such that $\chi \geq 0, \psi > 0$ if $\kappa > 0$; $\psi \geq 0, \chi > 0$ if $\kappa < 0$ and $\chi > 0, \psi > 0$ if $\kappa = 0$. An important property of the random variable X is that it can be written as a function of normal distribution via the normal mean-variance mixture representation. i.e., X admits the following stochastic representation:

$$X = \mu + W\lambda + W^{1/2}Z, \tag{3}$$

where the random variable Z is normally distributed with mean zero and variance σ^2, $\text{N}(0,\sigma^2)$, and W is a positive random variable, independent of Z, followed by the generalized inverse Gaussian (GIG) distribution with the pdf as

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

We write $W \sim \text{GIG}(\kappa,\chi,\psi)$ to indicate that the random variable W follows GIG distribution with parameters (κ,χ,ψ). A comprehensive survey of the GH and GIG distributions can be found in [33]. Using the law of iterative expectations, the mean and variance of X can be respectively obtained as

$$\begin{aligned} E(X) =& \mu + \lambda\left(\frac{\chi}{\psi}\right)^{\frac{1}{2}}R_{(\kappa,1)}(\sqrt{\chi\psi}),\\ Var(X) =& \left(\frac{\chi}{\psi}\right)^{\frac{1}{2}}\sigma^2 R_{(\kappa,1)}(\sqrt{\chi\psi}) + \lambda^2\Big[\left(\frac{\chi}{\psi}\right)R_{(\kappa,2)}(\sqrt{\chi\psi}) \\ & - \left(\left(\frac{\chi}{\psi}\right)^{\frac{1}{2}}R_{(\kappa,1)}(\sqrt{\chi\psi})\right)^2\Big], \end{aligned}$$

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

The Lindley distribution was originally proposed by Lindley [27] as an alternative distribution to model failure times

and has recently found several applications. A random variable W is said to have a Lindley distribution, denoted by Lindley(τ), if it is specified by the pdf

$$f_{L}(w;\tau) = \frac{\tau^2}{\tau+1}(w+1)e^{-\tau w}, \quad w>0,\ \tau>0,$$

where τ is a shape parameter. It follows immediately that the pdf of W can be formulated by a mixture of the exponential and gamma distributions with a mixing proportion $\tau/(\tau+1)$.

Consider Lindley(τ) as a mixing variable for (3). This gives rise to the NMVL distribution introduced by Naderi et al. [38], denoted by $X \sim \text{NMVL}(\mu,\lambda,\sigma^2,\tau)$, and the pdf and corresponding cumulative distribution function (cdf) of which are

$$\begin{aligned} f_{\text{NMVL}}(x;\mu,\lambda,\sigma^2,\tau) =& \frac{\tau}{1+\tau} f_{\text{GH}}(x;\mu,\lambda,\sigma^2,1,0,2\tau) \\ &+ \frac{1}{1+\tau} f_{\text{GH}}(x;\mu,\lambda,\sigma^2,2,0,2\tau), \\ F_{\text{NMVL}}(x;\mu,\lambda,\sigma^2,\tau) =& \frac{\tau}{1+\tau} F_{\text{GH}}(x;\mu,\lambda,\sigma^2,1,0,2\tau) \\ &+ \frac{1}{1+\tau} F_{\text{GH}}(x;\mu,\lambda,\sigma^2,2,0,2\tau), \end{aligned}$$

where $F_{\text{GH}}(\cdot)$ represents the cdf of GH distribution. The notation $X \sim \text{NMVL}(\mu,\lambda,\sigma^2,\tau)$ will be used if X followed by the NMVL distribution.

3. THE NMVL-BS DISTRIBUTION

If $Y \sim \text{NMVL}(0,\lambda,1,\tau)$ in (1), then we have a random variable T following the NMVL-BS distribution. Readily observed from (1) and (3), T can be generated by the following stochastic representation

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

where $Z \sim \text{N}(0,1)$, and independent of Z, $W \sim \text{Lindley}(\tau)$. This leads to obtain the cdf and pdf of T, respectively, as

(5)
$$\begin{aligned} f_{\text{NMVL-BS}}(t;\alpha,\beta,\lambda,\tau) =& A(t,\alpha,\beta) f_{\text{NMVL}}(a(t,\alpha,\beta);0,\lambda,1,\tau), \\ F_{\text{NMVL-BS}}(t;\alpha,\beta,\lambda,\tau) =& F_{\text{NMVL}}(a(t,\alpha,\beta);0,\lambda,1,\tau), \end{aligned}$$

where $t,\alpha,\beta,\tau>0$, $\lambda\in\mathbb{R}$ and $a(t,\alpha,\beta)$ and $A(t,\alpha,\beta)$ are defined in (2). The notation NMVL-BS($\alpha,\beta,\lambda,\tau$) will be used if T adopts the pdf (cdf) (5). Figure 1 displays some graphical representations of (5). It is clear that the density of NMVL-BS can take various shapes depending on its parameters. We can also see that the pdf (5) is both decreasing and skewed function. The curves on Figure 1 show that the skewness of NMVL-BS distribution is increased by increasing both α and λ. Moreover, for fixed α and λ, the skewness

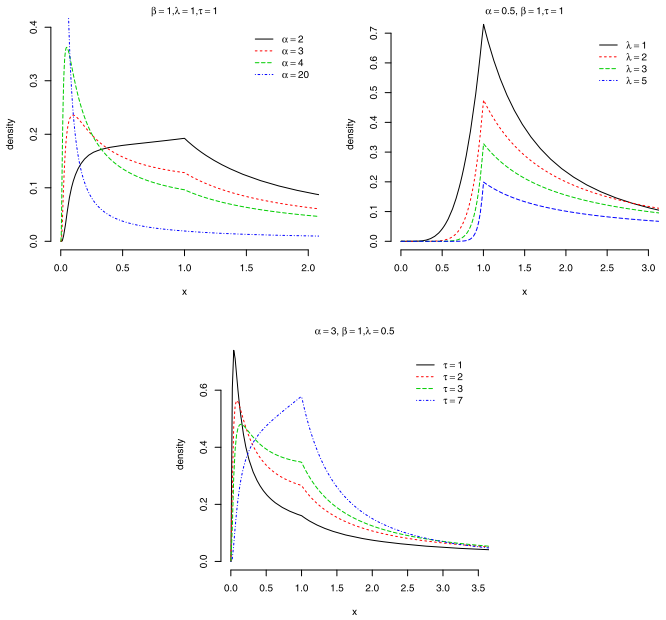

Figure 1. The density plots of NMVL-BS distribution for various parameters.

tends to negative values as τ gets large values. Therefore, the NMVL-BS distribution is flexible and the parameters λ and τ have substantial effects on its skewness and kurtosis. To illustrate the effects of shape parameters on the skewness and kurtosis, we obtain the skewness and kurtosis of T, respectively as

$$\gamma_t = \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3}{(\mu_2-\mu_1^2)^{1.5}}$$
$$\text{and} \quad \kappa_t = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4}{(\mu_2-\mu_1^2)^2},$$

where $\mu_r = E(T^r)$ for $r=1,2,3,4$. The closed form of μ_r are given in Appendix **A**.

From (5), the hazard rate function of NMVL-BS distribution is also given by

$$H(t) = \frac{f_{\text{NMVL-BS}}(t;\alpha,\beta,\lambda,\tau)}{1-F_{\text{NMVL-BS}}(t;\alpha,\beta,\lambda,\tau)}.$$

Table 1 provides numerical values of γ_t and κ_t for various value of shape parameters. It can be observed from Table 1 that the NMVL-BS distribution takes wide ranges of skewness and kurtosis. Moreover, the NMVL-BS distribution has negative skewness for both small values of α and negative values of λ. The plots of hazard rate function for some parameter values are drawn in Figure 2. Since the parameter β is a scale parameter, it does not change the shape of the hazard rate function. It is evident that the hazard rate function of the NMVL-BS distribution can be decreasing, increasing, upside-down bathtub (when λ

Table 1. Value of skewness and kurtosis based on moments of the NMVL-BS$(\alpha, \beta, \lambda, \tau)$ distribution when $\beta = 1$

		$\tau = 1$				$\tau = 2$			
		γ_t		κ_t		γ_t		κ_t	
$\lvert\lambda\rvert$	α	$-\lambda$	λ	$-\lambda$	λ	$-\lambda$	λ	$-\lambda$	λ
0.00	0.05	0.3995	0.3995	2.6837	2.6837	0.2845	0.2845	2.7865	2.7865
	0.10	0.8062	0.8062	3.7700	3.7700	0.5703	0.5703	3.3539	3.3539
	0.25	2.0902	2.0902	11.9450	11.9450	1.4646	1.4646	7.4516	7.4516
	0.50	4.0936	4.0936	37.1941	37.1941	3.0183	3.0183	22.6399	22.6399
	0.75	5.4547	5.4547	61.1426	61.1426	4.3905	4.3905	43.1644	43.1644
	1.00	6.2644	6.2644	77.5102	77.5102	5.4232	5.4232	62.1806	62.1806
0.05	0.05	0.2550	0.5435	2.5026	2.8898	0.1767	0.3921	2.6908	2.8963
	0.10	0.6565	0.9549	3.3723	4.1987	0.4605	0.6797	3.1492	3.5745
	0.25	1.9201	2.2564	10.6925	13.2330	1.3438	1.5843	6.8317	8.0965
	0.50	3.9476	4.2279	35.2593	38.9944	2.8867	3.1462	21.2279	24.0513
	0.75	5.3878	5.5098	60.2986	61.7788	4.2794	4.4952	41.5670	44.6853
	1.00	6.2695	6.2519	78.1662	76.7267	5.3505	5.4891	61.0748	63.1686
0.20	0.05	-0.1601	0.9525	2.1168	3.6260	-0.1396	0.7061	2.4905	3.3016
	0.10	0.2229	1.3759	2.3943	5.6210	0.1373	0.9984	2.6379	4.3175
	0.25	1.4062	2.7163	7.2829	17.1688	0.9833	1.9307	5.1488	10.1467
	0.50	3.4477	4.5595	28.8404	43.5043	2.4765	3.5034	17.0761	28.2019
	0.75	5.1092	5.6146	56.3830	62.6369	3.9114	4.7712	36.3894	48.7463
	1.00	6.2280	6.1812	79.0684	73.8809	5.0910	5.6494	57.0365	65.4672
0.50	0.05	-0.7726	1.5631	1.8692	5.2507	-0.6805	1.2409	2.4149	4.3158
	0.10	-0.4285	2.0191	1.3262	8.6063	-0.4197	1.5435	2.0496	5.9986
	0.25	0.5467	3.4155	2.6714	24.4193	0.3384	2.5289	2.6935	14.4664
	0.50	2.3333	4.9500	15.8644	48.7758	1.6604	4.0804	9.8808	35.6074
	0.75	4.2121	5.6437	42.6332	61.4332	3.0676	5.1629	25.2957	54.5709
	1.00	5.8000	5.9722	73.6523	67.6893	4.3935	5.8321	46.0488	67.6233
1.00	0.05	-1.1283	2.0395	1.7124	7.3752	-1.1921	1.7779	2.6053	5.8988
	0.10	-0.7989	2.6085	0.6475	12.9978	-0.9446	2.1186	1.7353	8.6436
	0.25	-0.0472	4.0831	0.0612	33.1348	-0.3071	3.2252	0.7775	21.2636
	0.50	1.0867	5.1617	3.7762	51.1508	0.6688	4.6842	2.9161	44.3560
	0.75	2.5035	5.5138	17.3809	57.3694	1.7581	5.4804	10.5157	59.0625
	1.00	4.2032	5.6564	44.7300	59.9051	2.9877	5.8990	25.0997	67.2161

is positive), increasing-decreasing-increasing or decreasing-increasing-decreasing (as λ tends to negative values). As a result, the new extended BS distribution is quite flexible and can be used effectively in analyzing left or right skewed lifetime data that have a monotone as well as non-monotone failure rates.

4. GENERAL PROPERTIES

In this section, some properties of the NMVL-BS distribution are studied. As a first result, the following theorem represents a hierarchical stochastic representation of NMVL-BS distribution which is useful for implementing the ECM algorithm discussed in the next section.

Theorem 4.1. *Let $T \sim$ NMVL-BS$(\alpha, \beta, \lambda, \tau)$ and $W \sim$ Lindley(τ), then the distribution of T given $W = w$ is*

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

where EBS denotes the extended BS distribution introduced by Leiva et al. [25].

Proof. The proof is straightforward through using (4) and the definition of EBS distribution. □

Proposition 4.2. *Let a random variable $T \sim$ NMVL-BS$(\alpha, \beta, \lambda, \tau)$. Then:*

i) $T^{-1} \sim$ NMVL-BS$(\alpha, \beta^{-1}, \lambda, \tau)$.
ii) $cT \sim$ NMVL-BS$(\alpha, c\beta, \lambda, \tau)$, *for any constant* $c > 0$.
iii) It can be shown that

$$X = \frac{1}{\alpha}\left[\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}}\right] \sim NMVL(0, \lambda, 1, \tau).$$

iv) The random variable T belongs to the class of scale mixture of normal Birnbaum-Saunders distributions [4], as $\lambda \to 0$.

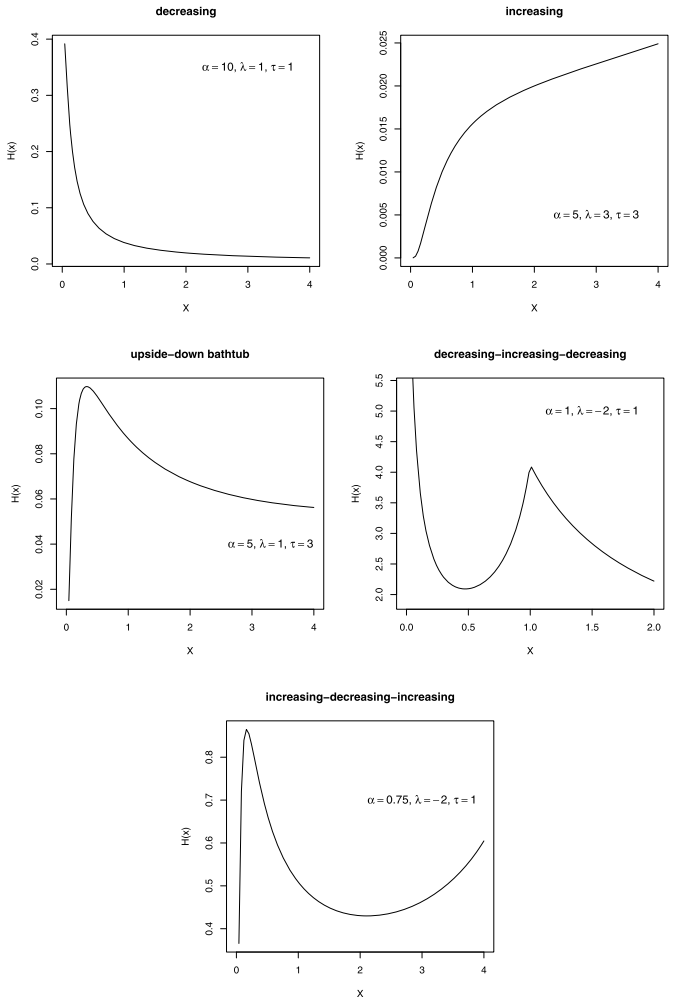

Figure 2. The plot of NMVL-BS hazard rate function for some parameter values; $\beta = 5$.

Proof. Parts (i) and (ii) are directly obtained from the change-of-variable method. The other parts can be easily proved by the mathematical works. □

The NMVL-BS quantile function can be easily determined from part (*iii*) of Proposition (4.2) as

(6)
$$Q_T(u) = \frac{\beta}{4}\left[\alpha F^{-1}(u) + \sqrt{\alpha^2 F^{-1}(u)^2 + 4}\right]^2, \quad u \in (0,1),$$

where $F^{-1}(u)$ is the quantile function of NMVL$(0, \lambda, 1, \tau)$. It is interesting to point out that the quantile function (6) can be used as a good alternative to generate a random sample from NMVL-BS$(\alpha, \beta, \lambda, \tau)$. It is also possible to investigate the effects of the additional shape parameters λ and τ on the skewness and kurtosis of the NMVL-BS distribution.

It is well-known that the classical kurtosis measure may tends to infinity when the underline distribution is heavy-tailed. Thus, the classical kurtosis measure becomes uninformative precisely when it needs to be. To deal with the infinity problem of asymmetry measure, the Bowley skewness ($\mathcal{B}$) [21] and the Moors kurtosis ($\mathcal{M}$) [37] are one of the earliest skewness and kurtosis measures which also exist even for distributions without moments. The Bowley skewness and the Moors kurtosis measures are less sensitive to outliers. In addition, $\mathcal{M}$ is not sensitive to variations of the values in the tails or to variations of the values around the median since it is defined based on the quantiles. For the NMVL-BS distribution, $\mathcal{B}$ and $\mathcal{M}$ are respectively defined as

$$\mathcal{B} = \frac{Q_T(3/4) + Q_T(1/4) - 2Q_T(1/2)}{Q_T(3/4) - Q_T(1/4)},$$
$$\mathcal{M} = \frac{Q_T(3/8) - Q_T(1/8) + Q_T(7/8) - Q_T(5/8)}{Q_T(3/4) - Q_T(2/8)}.$$

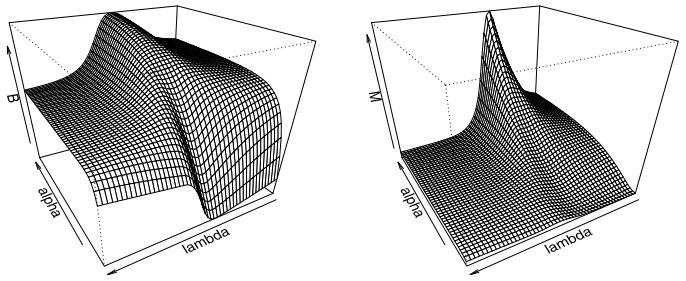

Figure 3. Plots of the Bowley skewness ($\mathcal{B}$) the Moors kurtosis ($\mathcal{M}$) as functions of α and λ for $\beta = 1, \tau = 0.5$.

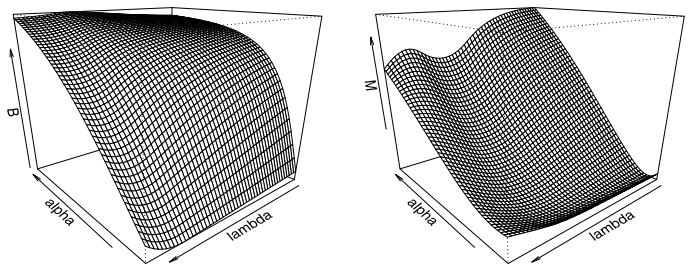

Figure 4. Plots of the Bowley skewness ($\mathcal{B}$) the Moors kurtosis ($\mathcal{M}$) as functions of α and λ for $\beta = 1, \tau = 1.5$.

Figures 3 and 4 demonstrate the form of measures $\mathcal{B}$ and $\mathcal{M}$ as functions of three shape parameters. Form the quantile function (6), it is clear that β has not an effect on the measures $\mathcal{B}$ and $\mathcal{M}$. The plots of $\mathcal{B}$ and $\mathcal{M}$ reveal that the skewness and kurtosis of the proposed distribution are substantially depend on the all of the shape parameters.

In the following theorem, we present a transformation result for the random variable T with NMVL-BS distribution which can be used to construct a linear regression model by assuming the NMVL-BS distribution for the unobserved error terms.

Theorem 4.3. *Let $T \sim$ NMVL-BS$(\alpha, \beta, \lambda, \tau)$. Then the random variable $Y = \log(T)$, called the Log-NMVL*

Birnbaum-Saunders distribution, has a following pdf:

$$f_Y(y;\alpha,\lambda,\tau,\mu) = U(y,\alpha,\mu) f_{NMVL}(u(y,\alpha,\mu);0,\lambda,1,\tau),$$

where $\mu = \log(\beta)$,

$$U(y,\alpha,\mu) = \frac{2}{\alpha}\cosh\left(\frac{y-\mu}{2}\right),$$
$$\text{and } u(y,\alpha,\mu) = \frac{2}{\alpha}\sinh\left(\frac{y-\mu}{2}\right).$$

Proof. The theorem is proved through using the change-of-variable method. □

Consequently, the following theorem is established, which is useful for the calculation of some conditional expectations involved in the proposed ECM algorithm.

Theorem 4.4. *Let* $W \sim$ *Lindley*(τ) *and* $T \sim$ *NMVL-BS*$(\alpha,\beta,\lambda,\tau)$. *Then, the conditional distribution of* W *given* $T = t$, *obtained by Bayes' rule, is*

$$f_{W|T=t}(w;\mu,\lambda,\sigma^2,\tau) = \pi(t) f_{GIG}(w;0.5,\chi^*,\psi^*) + (1-\pi(t)) f_{GIG}(w;1.5,\chi^*,\psi^*),\ w>0,$$

where $\chi^* = a^2(t,\alpha,\beta)$, $\psi^* = 2\tau + \lambda^2$, *and*

$$\pi(t) = \tau f_{GH}(a(t,\alpha,\beta);0,\lambda,1,1,0,2\tau) \Big/ \Big(\tau f_{GH}(a(t,\alpha,\beta);0,\lambda,1,1,0,2\tau) + f_{GH}(a(t,\alpha,\beta);0,\lambda,1,2,0,2\tau)\Big).$$

Moreover, for $r = \pm1, \pm2, \ldots,$

$$E(W^r|T=t) = \left(\frac{\chi^*}{\psi^*}\right)^{r/2}\left(\pi(t) R_{(0.5,r)}\left(\sqrt{\chi^*\psi^*}\right) + (1-\pi(t)) R_{(1.5,r)}\left(\sqrt{\chi^*\psi^*}\right)\right). \quad (7)$$

5. PARAMETER ESTIMATION VIA ECM ALGORITHM

In this section, we demonstrate how to carry out the ECM algorithm to compute ML estimation of the proposed NMVL-BS distribution. The ECM algorithm is a commonly used optimization strategy to obtain ML estimates in incomplete data problems. This iterative algorithm is a variant of the EM algorithm with the maximization (M) step of EM replaced by a sequence of computationally simpler conditional maximization (CM) steps.

Consider n independent random variables $\boldsymbol{T} = (T_1,\ldots,T_n)^\top$ which are followed by a NMVL-BS distribution with unknown parameter vector $\boldsymbol{\theta} = (\alpha,\beta,\lambda,\tau)$. So, the log-likelihood function for a given set of observed data $\boldsymbol{t} = (t_1,\ldots,t_n)^\top$ is given by

$$\ell(\boldsymbol{\theta}|\boldsymbol{t}) = \sum_{i=1}^n \log f_{\text{NMVL-BS}}(t_i;\alpha,\beta,\lambda,\tau). \quad (8)$$

Although the ML estimator of $\boldsymbol{\theta}$ is theoretically obtained by solving $\boldsymbol{\theta} = \arg\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}|\boldsymbol{t})$, a direct maximization of (8) can be computationally problematic, since its derivatives with respect to parameters are difficult to compute. In order to apply the ECM algorithm in the NMVL-BS distribution, the hierarchical representation of (8) obtained by theorem 4.1 is

$$\begin{aligned} T_i\big|W_i = w_i &\sim \text{EBS}(\alpha\sqrt{w_i},\beta,2,-\lambda\sqrt{w_i},0), \\ W_i &\sim \text{Lindley}(\tau). \end{aligned} \quad (9)$$

The hierarchical representation (9) leads to the complete-data log-likelihood function for $\boldsymbol{\theta} = (\alpha,\beta,\lambda,\tau)$ associated with the observed responses $\boldsymbol{t} = (t_1,\ldots,t_n)^\top$ and hidden variables $\boldsymbol{w} = (w_1,\ldots,w_n)^\top$, omitting additive constants, as

$$\begin{aligned} \ell_c(\boldsymbol{\theta}) &= \sum_{i=1}^n \log f_{T|W}(t_i|w_i,\alpha,\beta,\lambda) + \sum_{i=1}^n \log f_W(w_i;\tau) \\ &= \ell_{c1}(\alpha,\beta,\lambda|\mathbf{t},\mathbf{w}) + \ell_{c2}(\tau|\mathbf{w}), \end{aligned} \quad (10)$$

where $\ell_{c2}(\tau|\mathbf{w}) = n\log\left(\frac{\tau^2}{1+\tau}\right) - \tau\sum_{i=1}^n w_i$, and

$$\ell_{c1}(\alpha,\beta,\lambda \mid \mathbf{t},\mathbf{w}) = -n\log(\alpha) - \frac{n}{2}\log(\beta) + \sum_{i=1}^n \log(t_i+\beta) - \frac{1}{2\sigma^2}\sum_{i=1}^n \frac{1}{w_i}\left(\frac{t_i}{\beta} + \frac{\beta}{t_i} - 2\right) - \frac{\lambda^2}{2}\sum_{i=1}^n w_i + \frac{\lambda}{\alpha}\sum_{i=1}^n \xi(t_i,\beta),$$

in which $\xi(t,\beta) = a(t,1,\beta)$.

The ECM algorithm for ML estimation of the NMVL-BS model can be finally summarized through the following two iterative steps.

- **E-step:** At iteration k, the so-called Q-function, defined as the conditional expectation of the complete data log-likelihood function (10) evaluated at $\hat{\boldsymbol{\theta}}^{(k)}$, is computed as

$$Q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}^{(k)}) = Q_1(\alpha,\beta,\lambda \mid \hat{\boldsymbol{\theta}}^{(k)}) + Q_2(\tau \mid \hat{\boldsymbol{\theta}}^{(k)}), \quad (11)$$

  where for the conditional expectations $\hat{w}_i^{(k)} = E\left(W_i|T_i = t_i, \hat{\boldsymbol{\theta}}^{(k)}\right)$, and $\hat{u}_i^{(k)} = E\left(W_i^{-1}|T_i = t_i, \hat{\boldsymbol{\theta}}^{(k)}\right)$ calculated by using (7) in Proposition 4.2,

$$Q_1\left(\alpha,\beta,\lambda \mid \hat{\boldsymbol{\theta}}^{(k)}\right) = -n\log(\alpha) - \frac{n}{2}\log(\beta) + \sum_{i=1}^n \log(t_i+\beta)$$

$$
\begin{aligned}
&- \frac{1}{2\alpha^2}\sum_{i=1}^{n}\hat{u}_i^{(k)}\left(\frac{t_i}{\beta}+\frac{\beta}{t_i}-2\right)\\
&- \frac{\lambda^2}{2}\sum_{i=1}^{n}\hat{w}_i^{(k)}+\frac{\lambda}{\alpha}\sum_{i=1}^{n}\xi(t_i,\beta),\\
Q_2\big(\tau \mid \hat{\boldsymbol{\theta}}^{(k)}\big) =& n\log\left(\frac{\tau^2}{1+\tau}\right)-\tau\sum_{i=1}^{n}\hat{w}_i^{(k)}.
\end{aligned}
$$

- **CM-steps:** Let $S_u^{(k)} = \frac{1}{n}\sum_{i=1}^{n} t_i\hat{u}_i^{(k)}$, $R_u^{(k)} = \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^{(k)}/t_i$, $\bar{U}^{(k)} = \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^{(k)}$, $\bar{w}^{(k)} = \frac{1}{n}\sum_{i=1}^{n}\hat{w}_i^{(k)}$, and $\bar{\xi}^{(k)} = \frac{1}{n}\sum_{i=1}^{n}\xi(t_i,\hat{\beta}^{(k)})$. For updating $\hat{\boldsymbol{\theta}}^{(k+1)}$, maximize (11) over α,λ and τ. This leads to the following CM estimators:

$$
\begin{aligned}
\hat{\alpha}^{2(k+1)} &= \frac{S_u^{(k)}}{\hat{\beta}^{(k)}}+\hat{\beta}^{(k)}R_u^{(k)}-2\bar{U}-\frac{\bar{\xi}^{2(k)}}{\bar{w}^{(k)}},\\
\hat{\tau}^{(k+1)} &= \frac{(1-\bar{w}^{(k)})+\sqrt{(\bar{w}-1)^2+8\bar{w}^{(k)}}}{2\bar{w}^{(k)}},\\
\hat{\lambda}^{(k+1)} &= \frac{\bar{\xi}^{(k)}}{\hat{\alpha}^{(k+1)}\bar{w}^{(k)}}.
\end{aligned}
$$

  Since there is no closed-form solution for $\hat{\beta}^{(k+1)}$, we update β by maximizing the constrained actual observed Q-function. So, we have

$$
\hat{\beta}^{(k+1)} = \arg\max_{\beta} Q_1(\hat{\alpha}^{(k+1)},\beta,\hat{\lambda}^{(k+1)}|\boldsymbol{t},\boldsymbol{w}^{(k)}).
$$

The above procedure is iterated until a suitable convergence rule is satisfied, e.g. $\big|\ell(\hat{\boldsymbol{\theta}}^{(k+1)})-\ell(\hat{\boldsymbol{\theta}}^{(k)})\big|$ is less than a user-specified tolerance where $\ell(\hat{\boldsymbol{\theta}}^{(k)})$ is the maximized log-likelihood function at iteration k. As an alternative stopping criterion, we shall use the Aitken acceleration [2] method to avoid an indication of lack of progress of the algorithm [34]. Following [10, 28], the ECM algorithm can be considered to have reached convergence when $\ell_\infty(\hat{\boldsymbol{\theta}}^{(k+1)})-\ell(\hat{\boldsymbol{\theta}}^{(k)})<\varepsilon$, where the asymptotic estimate of the log-likelihood at iteration $k+1$ is

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

in which

$$
a^{(k)} = \frac{\ell(\hat{\boldsymbol{\theta}}^{(k+1)})-\ell(\hat{\boldsymbol{\theta}}^{(k)})}{\ell(\hat{\boldsymbol{\theta}}^{(k)})-\ell(\hat{\boldsymbol{\theta}}^{(k-1)})}.
$$

In our study, the tolerance ε is considered as 10^{-5}.

The EM algorithm may not give maximum global solution as the initial value $(\hat{\Theta}^{(0)})$ is far from the real parameter value. Therefore, the chosen starting points play an important role in parameter estimation. We create the initial value $\hat{\alpha}^{(0)}$ and $\hat{\beta}^{(0)}$ by, for example, the modified moment estimates proposed by Ng et al. [39], and $\hat{\lambda}^{(0)}=0$ and $\hat{\tau}^{(0)}=1$ [38].

To compute the asymptotic covariance of the ML estimate, $\hat{\boldsymbol{\theta}}=(\hat{\alpha},\hat{\beta},\hat{\lambda},\hat{\tau})$, we use the information-based method suggested by Meilijson [35]. Formally, the empirical information matrix is defined as

$$
I_e(\boldsymbol{\theta}\mid\boldsymbol{t}) = \sum_{j=1}^{n}\boldsymbol{s}(t_j|\boldsymbol{\theta})\boldsymbol{s}^\top(t_j|\boldsymbol{\theta})-\frac{1}{n}\boldsymbol{S}(\mathbf{t}|\boldsymbol{\theta})\boldsymbol{S}^\top(\mathbf{t}|\boldsymbol{\theta}), \tag{12}
$$

where $\boldsymbol{S}(\boldsymbol{t}\mid\boldsymbol{\theta})=\sum_{j=1}^{n}\boldsymbol{s}(t_j|\boldsymbol{\theta})$, and $\boldsymbol{s}(t_j\mid\boldsymbol{\theta})$ are individual scores which can be determined from the result of Louis [29] as

$$
\boldsymbol{s}(t_j|\boldsymbol{\theta}) = \frac{\partial f(t_j|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = E\left(\frac{\partial\ell_c(\boldsymbol{\theta}\mid t_j,w_j)}{\partial\boldsymbol{\theta}}\Big|t_j,\boldsymbol{\theta}\right).
$$

Substituting the ML estimates $\hat{\boldsymbol{\theta}}$ into (12) gives

$$
I_e(\boldsymbol{\theta}\mid\boldsymbol{t}) = \sum_{j=1}^{n}\hat{\boldsymbol{s}}_j\hat{\boldsymbol{s}}_j^\top, \tag{13}
$$

where $\hat{\boldsymbol{s}}_j=(\hat{s}_{j,\alpha},\hat{s}_{j,\beta},\hat{s}_{j,\lambda},\hat{s}_{j,\tau})^\top$. The explicit expressions for the elements of $\hat{\boldsymbol{s}}_j$ are

$$
\begin{aligned}
\hat{s}_{j,\tau} =& \frac{2}{\hat{\tau}}-\frac{1}{1+\hat{\tau}}-\hat{w}_j,\\
\hat{s}_{j,\lambda} =& a(t_j,\hat{\alpha},\hat{\beta})-\hat{\lambda}\hat{w}_j,\\
\hat{s}_{j,\alpha} =& \frac{\hat{u}_j}{\hat{\alpha}^3}\Big(\frac{t_j}{\hat{\beta}}+\frac{\hat{\beta}}{t_j}-2\Big)-\frac{1}{\hat{\alpha}}-\frac{\hat{\lambda}}{\hat{\alpha}}a(t_j,\hat{\alpha},\hat{\beta}),\\
\hat{s}_{j,\beta} =& \frac{1}{t_j+\hat{\beta}}-\frac{1}{2\hat{\beta}}-\frac{\hat{u}_j}{2\hat{\alpha}^2}\left(\frac{1}{t_j}-\frac{t_j}{\hat{\beta}}\right)\\
&-\frac{\hat{\lambda}}{2\hat{\alpha}\hat{\beta}}\left(\sqrt{t_j/\hat{\beta}}+\sqrt{\hat{\beta}/t_j}\right).
\end{aligned}
$$

As a result, the standard errors of $\hat{\boldsymbol{\theta}}$ are taken as the square roots of the diagonal elements of the inverse of (13).

6. DATA ANALYSIS

In this section, the proposed model is applied to two real data for illustrative purposes. For the sake of model comparison, we also consider BS, SN-BS, ST-BS, SNT-BS and student-t-BS (T-BS) distributions. The models in competition are compared using the AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) and HQIC (Hannan-Quinn Information Criterion), criteria defined as

$$
\begin{aligned}
&\mathrm{AIC} = 2m-2\ell_{\max},\quad \mathrm{BIC} = m\log n-2\ell_{\max},\\
&\mathrm{HQIC} = m\log\log n-2\ell_{\max},
\end{aligned}
$$

where m and n are the number of free parameters and sample size, respectively, and $\ell_{\max}$ is the maximized log-likelihood value. It should be note that models with lower

Table 2. ML estimates and standard errors of the fitted models on S1

	BS		SN-BS		T-BS		ST-BS		SNT-BS		NMVL-BS	
parameter	MLE	SE	MLE	SE	MLE	SE	MLE	SE	MLE	SE	MLE	SE
α	0.565	0.073	0.619	0.107	0.431	0.096	0.451	0.145	0.654	0.094	0.855	0.163
β	8.042	0.827	6.395	1.546	8.978	0.928	14.515	1.436	5.614	0.216	14.180	0.027
λ			0.558	0.437			-2.130	1.213	51.562	8.955	-4.261	0.447
ν					4.301	3.455	2.011	0.829	0.221	0.311		
τ											7.172	1.124
$\ell_{\max}$	-96.655		-96.265		-95.728		-92.851		-94.463		-90.855	
AIC	197.309		198.532		197.456		193.702		196.926		189.709	
BIC	200.303		203.021		201.946		199.688		202.912		195.695	
HQIC	195.813		196.287		195.212		190.709		193.933		186.716	
LR-Test	11.600		10.820		9.746		3.992		7.216			
$2\log(B_{12})$	4.607		7.323		6.249		3.992		7.216			
KS-Test	0.174		0.169		0.120		0.122		0.680		0.058	
P-value	0.245		0.263		0.682		0.674		< 0.001		0.989	

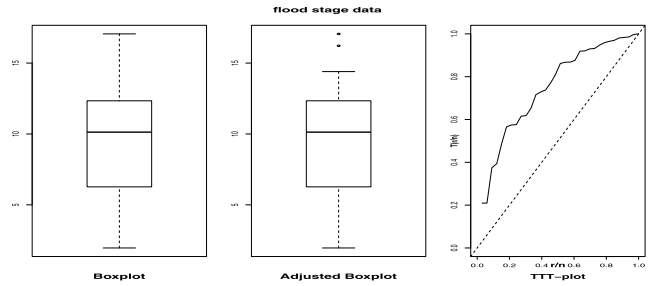

Figure 5. The standard box plot (left panel), adjusted box-plot (mid panel) and TTT plot (right panel) for the flood stage data.

values of AIC, BIC or HQIC are considered more preferable. We also apply formal goodness-of-fit test, Kolmogorov-Smirnov (K-S), in order to verify how well the models describes data. In general, the smaller the values of K-S statistics, the better the fit to the data. Moreover, the reported p-value of K-S test can be used as a similarity assessment of the experimental data against the fitted distribution.

6.1 Application on the flood stage data

The first data set (S1) considered here is on the flood stage for two stations on the Fox River in Wisconsin. The data were originally studied by Kappenman [20] and also analyzed by Barbiero [5].

The standard box, adjusted box and TTT (total time on test) plots for data are presented in Figure 5. The standard box plot indicates that there are no atypical observations on the both tails, however this plot was constructed for symmetric data. By drawing the adjusted box-plot, we can observe that there exist some atypical observations on the right-tail. It is also worthwhile to note that the TTT plot is a straight diagonal for constant failure rate, and convex (concave) for decreasing (increasing) ones. It is also first concave (convex) and then convex (concave) if the failure rate is upside-down bathtub (bathtub) shaped. The TTT plot of S1 demonstrates that these data seem to have a hazard rate function that is coherent with that of the distribution with a concave hazard rate function. Therefore, these results provide diverse evidences for proposing the NMVL-BS model in order to describe the flood stage data.

Table 2 lists the ML results obtained by fitting the six considered models. Results based on AIC, BIC and HQIC indicate that the NMVL-BS distribution provides an improved fit of the data over five competitors. Due to the effect of sample size on the mentioned measures, we use Bayes factor, B_{01}, and the standard likelihood ratio (LR) tests [45] in order to highlight differences between the fitted models. These tests are applied to assay the hypothesis H_0: data are given from NMVL-BS distribution against H_1: data have arisen under one of the five competitor distributions. Specifically, for given priori probabilities $\mathbb{P}_j(\boldsymbol{\theta})$ under H_j $(j = 0, 1)$, the factor B_{01} is defined by

$$B_{01} = \frac{\mathbb{P}(\boldsymbol{t}|H_0)}{\mathbb{P}(\boldsymbol{t}|H_1)} = \frac{\int \ell(\boldsymbol{t}|H_0)\mathbb{P}_0(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int \ell(\boldsymbol{t}|H_1)\mathbb{P}_1(\boldsymbol{\theta})d\boldsymbol{\theta}},$$

where $\ell(\boldsymbol{t}|H_j)$ for $j = 0, 1$, represents the log-likelihood under the model H_j. Following [40], the B_{01} can be approximated by

$$2\log(B_{01}) \approx LR - (m_0 - m_1)\log(n),$$

where $LR = 2(\ell(\boldsymbol{t}|H_1) - \ell(\boldsymbol{t}|H_0))$ is the standard likelihood ratio statistic and m_j is a number of free parameters under H_j. Using the results of LR and Bayes factor tests in Table 2 and interpretation of Bayes factor test described in Table 10 of Vilca et al. [45], we have a positive evidence to reject the alternative hypotheses in favor of the NMVL-BS distribution. It means that the NMVL-BS distribution is significantly better than the other distributions for flood stage data. The reported p-values of K-S test also suggests strongly that S1 follow a NMVL-BS distribution. Finally,

Table 3. ML estimates and standard errors of the fitted models on S2

	BS		SN-BS		T-BS		ST-BS		SNT-BS		NMVL-BS	
parameter	MLE	SE	MLE	SE	MLE	SE	MLE	SE	MLE	SE	MLE	SE
α	0.150	0.004	0.223	0.007	0.078	0.069	0.112	0.099	0.217	0.007	0.297	0.001
β	8.151	0.046	6.922	0.039	7.815	0.179	7.227	0.219	6.977	0.038	7.515	0.042
λ			3.577	0.303			2.295	0.555	7.838	1.810	1.628	0.023
ν					1.939	0.458	2.244	0.555	1.718	0.480		
τ											6.753	0.063
$\ell_{\max}$	-1133.024		-1054.357		-1066.965		-1013.124		-1042.363		-1004.121	
AIC	2270.048		2114.714		2139.930		2034.248		2092.726		2016.242	
BIC	2281.064		2131.239		2156.455		2056.281		2114.759		2038.275	
HQIC	2270.081		2114.762		2139.978		2034.312		2092.79		2016.306	

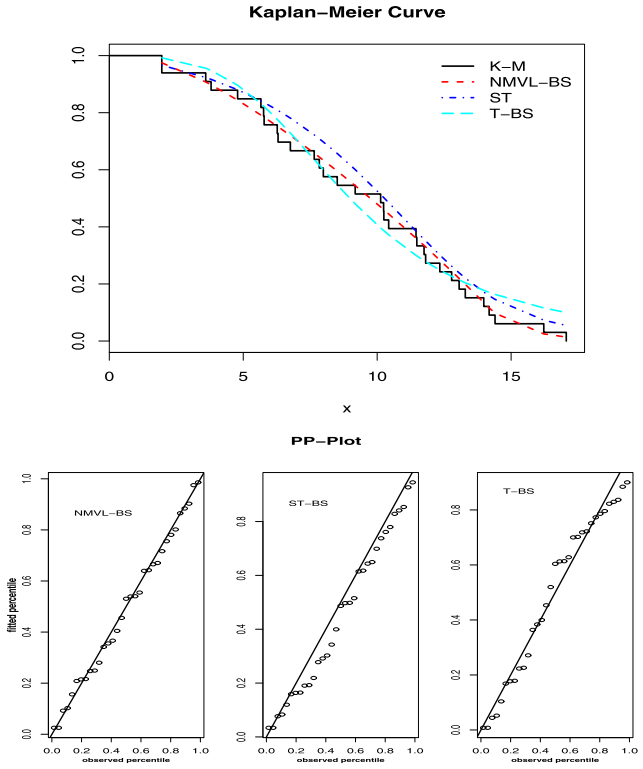

Figure 6. The Kaplan-Meier (top panel) curve and pp-plots for NMVL-BS, ST-BS and T-BS (down panel) of S1.

in order to highlight the outperformance of NMVL-BS distribution on S1, the Kaplan-Meier carve and *pp*-plots of the three best models are drawn in Figure 6. The results of these plots are consistent with those obtained through the AIC, BIC, HQIC, and Bayes factor, LR and K-S tests.

6.2 Application on the fire loss data

In the second experimental real analysis, we consider fire insurance claim data set, called beaonre in the `R` package **CASdatasets**, which concerns claim information from office buildings. The variable log-SumInsured (S2) with size 1823 is used for our illustrative purpose. We fit the six considered distribution on the data. Table 3 displays the ML results obtained by fitting the six considered models. Results based on AIC, BIC, and HQIC suggest that the NMVL-BS distribution provides the best fit for S2.

The evaluation of risk is a major subject for the investors who are holding a position on portfolios of risky assets. So, the risk measures and their theories play an important role in estimating financial losses. Among the several purposes of the risk measure, the most important ones in practice are: determination of risk capital and capital adequacy, management tool and insurance premiums [33]. To attain these purposes, statistical tools play a substantial role since most modern measures of the risk in a portfolio are statistical quantities. For the sake of model comparison in risk measure theory, we compute VaR and TVaR measures. The VaR is a widely risk measure used in capital markets. For a given level q, the VaR is defined as the q-quantile of the distribution. i.e. $VaR_q = \inf\{x|F(x) > q\}$. Since the VaR is not a coherence measure (satisfying the monotonicity, subadditivity, homogeneity, and translational invariance properties), an alternative risk measure, TVaR is defined that gives the expected amount of extreme loss under a given risk. Theoretically, the TVaR of a random variable X for a given level q is

$$\mathrm{TVaR}_q^X = E[X|X \geq VaR_q].$$

Based on the result of fitting distributions, we compare the six considered models in terms of its accuracy to predict VaR and TVaR. Figure 7 shows the empirical VaR and TVaR along with their predicted values obtained from the six BS-kind distributions with levels ranging between 50% and 100%. It can be seen that both NMVL-BS and ST-BS distributions predict the VaR and TVaR much better than the other models.

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

$$\mathrm{MARE}(\hat{M}) = \frac{1}{n_q}\sum_{i=1}^{n_q}\Big|\frac{M-\hat{M}_i}{M}\Big|,$$

Table 4. MARE of VaR and TVaR (in %)

Risk measures	BS	T-BS	SN-BS	ST-BS	SNT-BS	NMVL-BS
VaR	4.6408	3.6572	3.8417	2.6395	4.1031	1.3453
TVaR	3.2574	2.0749	2.1344	1.2395	2.5361	0.7105

Table 5. Simulation results for assessing robustness of the model on outlier with various sample sizes

noises	n	Criteria	BS	T-BS	SN-BS	SNT-BS	ST-BS	NMVL-BS
2%	250	AIC	1298.5	1294.1	1278.2	1254.9	1253.0	1247.4
		BIC	1305.6	1304.7	1288.8	1269.0	1267.2	1261.6
	500	AIC	2585.4	2575.3	2542.8	2496.0	2492.2	2481.1
		BIC	2593.8	2588.0	2555.5	2512.9	2509.1	2498.1
	1000	AIC	5161.6	5140.2	5072.7	4976.5	4967.6	4945.9
		BIC	5171.4	5155.0	5087.5	4996.2	4987.3	4965.6
	2000	AIC	10310.6	10269.8	10135.0	9947.1	9926.4	9886.0
		BIC	10321.8	10286.7	10151.9	9969.6	9948.9	9908.5
4%	250	AIC	1327.2	1323.3	1306.7	1283.2	1279.8	1275.4
		BIC	1334.3	1333.9	1317.4	1297.4	1294.0	1289.6
	500	AIC	2636.6	2627.6	2593.7	2546.5	2542.8	2530.8
		BIC	2645.1	2640.3	2606.5	2563.5	2559.8	2547.8
	1000	AIC	5264.0	5241.3	5173.1	5075.4	5057.8	5044.2
		BIC	5273.8	5256.2	5187.9	5095.2	5077.6	5063.9
	2000	AIC	10508.9	10464.5	10326.6	10132.7	10111.0	10071.4
		BIC	10520.0	10481.4	10343.5	10155.3	10130.8	10093.9

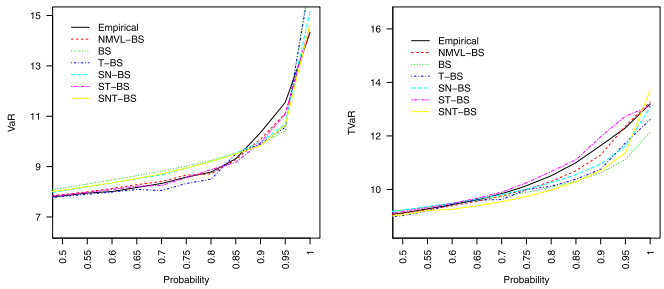

Figure 7. S2 data: model comparison in terms of VaR and TVaR.

where n_q represents the number of level considered on unit interval, M and $\hat{M}$ are the empirical and its corresponding predicted risk measures. The results summarized in Table 4 indicate that the NMVL-BS model outperforms the others in empirical estimation of VaR and TVaR since it has less amount of MARE.

7. SIMULATION ANALYSIS

In this section, two simulations are carried out to examine the performance of our model and its computational method. The first simulation has been done aims at showing the robustness on estimating the NMVL-BS distribution in which some levels of outlier are added to the simulated data. In the second simulation, we demonstrate if the proposed estimating approach can provide good asymptotic properties.

7.1 Robustness of the NMVL-BS model

In this experiment, 500 Monte Carlo samples from $BS(1,3)$ distribution are simulated for various sample size $n = 250, 500, 1000$, and 2000. In each replication, we also add some outlier points with two levels 2% and 4% which are generated uniformly on $[10, 20]$. Then, the parameter estimates are computed under BS, T-BS, SN-BS, SNT-BS, ST-BS and NMVL-BS distributions with two contaminations.

Table 5 summarizes the average of AIC and BIC. It can be seen that the influence of the outliers in model section criteria increases as the level of noisy points increase for all models. It is clear that the NMVL-BS model is less adversely affected by outliers, indicating that the proposed distribution is robust against the presence of outlier observations. On the other hand, it seems that an extreme observation is much more effective on the BS distribution, reflecting a lack

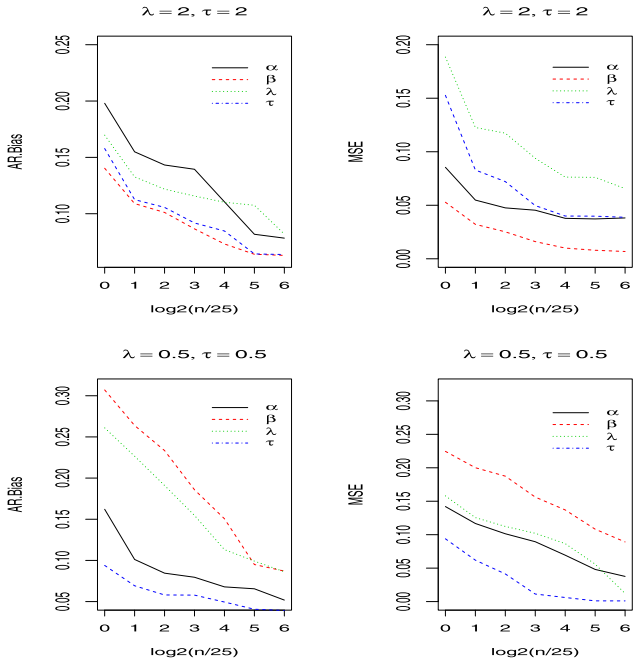

Figure 8. A comparison of AR. bias and MSE in parameter estimates for various sample sizes and parameters.

of enough ability to reduce the influence of outliers for BS model.

7.2 Finite sample properties of ML estimates

The second simulation is conducted to investigate the finite-sample properties of the ML parameter estimates. We generate synthetic samples of size $n = 25, 50, 100, 200, 400, 800$ and 1600 in each 500 replications from the NMVL-BS distribution with true considered parameters $\alpha = \beta = 1$ and $\lambda = \tau = 0.5, 2$. In order to examine the accuracies of the ML estimates, the absolute relative bias (AR. bias) and the root mean square error (MSE) are computed as

$$\text{AR.bias} = \frac{1}{500}\sum_{i=1}^{500}\left|\frac{\hat{\theta}_i - \theta}{\theta}\right| \quad \text{and}$$

$$\text{MSE} = \frac{1}{500}\sum_{i=1}^{500}(\hat{\theta}_i - \theta)^2 + \left[\frac{1}{500}\sum_{i=1}^{500}\hat{\theta}_i - \theta\right]^2.$$

where $\hat{\theta}_i$ denotes the parameter estimates of α, β, λ and τ obtained from the ith replication. Figures 8 displays a graphical representation of the AR. bias and the MSE of the four parameter estimates as a function of sample size n. Clearly, both magnitudes of AR. bias and MSE values converge reasonably well toward zero by increasing the sample size. It shows that the ML estimates obtained via the ECM algorithm are empirically consistent.

8. CONCLUSION

In this paper, we have dealt with the new extension of BS distribution through considering the NMVL distribution in (1). The proposed model, called the NMVL-BS distribution, has one scale parameter and three shape parameters which can takes various forms depending on them. We have showed that the NMVL-BS distribution is a positive as well as negative skewed distribution that can take wide range of skewness and kurtosis. It is interesting to point out that the hazard rate function of NMVL-BS distribution can takes decreasing, increasing, upside-down bathtub, increasing-decreasing-increasing or decreasing-increasing-decreasing forms allow us to use it in practical analysis. Some properties of the NMVL-BS distribution have been studied and a convenient hierarchical representation for the NMVL-BS distribution has presented. The feasible ECM algorithm is developed for computing maximum likelihood estimates. Numerical results suggest that the proposed NMVL-BS distribution can be served as an alternative model to the BS, T-BS, SN-BS, ST-BS and SNT-BS distributions for modeling positive experimental data.

The utility of current work can be extended to explore the multivariate case [26]. In addition, the finite mixture model based on the NMVL-BS distribution can be constructed through applying the k-bumps algorithm as an initialization strategy in the EM algorithm [8].

APPENDIX A

Let $T \sim \text{NMVL-BS}(\alpha, \beta, \lambda, \tau)$ and $Y \sim \text{NMVL}(0, \lambda, 1, \tau)$. In order to calculate skewness and kurtosis of T, by (4) and simple mathematical work, we have

$$\begin{aligned}
E(T) =& \frac{1}{2}\beta\alpha^2 E(Y^2) + 1 + \frac{1}{2}\alpha\beta V_1,\\
E(T^2) =& \frac{1}{2}\beta^2\alpha^4 E(Y^4) + 1 + \alpha^2\beta(1+\beta)E(Y^2)\\
&+ \alpha\beta V_1 + \frac{1}{2}\alpha^3\beta^2 V_3,\\
E(T^3) =& 1 + \frac{1}{2}\beta^3\alpha^6 E(Y^6) + \frac{1}{2}\alpha^5\beta^3 V_5 + \frac{3}{2}\alpha^4\beta^2(\beta+1)E(Y^4)\\
&+ \frac{3}{2}\alpha^3\beta^2(\beta+1)V_3 + 3\alpha^2\beta(\beta+\frac{3}{2})E(Y^2) + \frac{3}{2}\alpha\beta V_1,\\
E(T^4) =& 1 + \frac{1}{4}\beta^4\alpha^8 E(Y^8) + \frac{1}{2}\beta^4\alpha^7 V_7 + \beta^3\alpha^6(\frac{3}{2}+\beta)E(Y^6)\\
&+ \beta^3\alpha^5(2+\beta)V_5 + \beta^2\alpha^4(3+4\beta+\beta^2)E(Y^4)\\
&+ 2\alpha^3\beta^2(1+\beta)V_3 + 6\alpha^2\beta^2 E(Y^2),
\end{aligned}$$

where $V_r = E(Y^r\sqrt{\alpha^2 Y^2 + 4})$, for $r = 1, 3, 5, 7$ which are calculated numerically. Furthermore, since $Y|W = w \sim N(\lambda w, w)$, we have

$$\begin{aligned}
E(Y^2) =& \lambda^2 E(W^2) + E(W),\\
E(Y^4) =& \lambda^4 E(W^4) + 6\lambda^2 E(W^3) + 3E(W^2),
\end{aligned}$$

$$E(Y^6) = \lambda^6 E(W^6) + 15\lambda^4 E(W^5) + 45\lambda^2 E(W^4) + 15E(W^3),$$
$$E(Y^8) = \lambda^8 E(W^8) + 28\lambda^6 E(W^7) + 210\lambda^4 E(W^6) + 420\lambda^2 E(W^5) + 105E(W^4),$$

where $E(W^r) = \frac{r!(\alpha+r+1)}{\alpha^r(\alpha+1)}$.

ACKNOWLEDGMENTS

The authors deeply thank the associated editor and two anonymous referees for their valuable suggestions, corrections and encouragement, which have improved considerably earlier versions of the manuscript. M. Naderi's work is based upon research supported by the National Research Foundation, South Africa (Reference: CPRR160403161466 Grant Number: 105840 and STATOMET).

All the graphical and numerical computations presented here have been performed using the R language and the code is available from the first author upon request.

Received 13 September 2018

REFERENCES

[1] Aas, K. (2006). The generalized hyperbolic skew student's t distribution. *J. Financ. Econometrics* **4** 275–309.

[2] Aitken, A. C. (1927). On Bernoulli's numerical solution of algebraic equations. *Proc. Roy. Soc. Edinb.* **46** 289–305.

[3] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Stat.* **12** 171–178. MR0808153

[4] Balakrishnan, N., Leiva, V., Sanhueza, A. and Vilca, F. (2009). Estimation in the Birnbaum-Saunders distribution based on scale-mixture of normals and the EM-algorithm. *Sort* **33** 171–192. MR2643505

[5] Barbiero, A. (2013). An alternative discrete skew Laplace distribution. *J. Stat. Methodol.* **16** 47–67. MR3110887

[6] Barros, M., Galea, M., Leiva, V. and Santos-Neto, M. (2018). Generalized Tobit models: diagnostics and application in econometrics. *J. Appl. Stat.* **45** 145–167. MR3736863

[7] Bee, M., Dickson, M. M. and Santi, F. (2017). Likelihood-based risk estimation for variance-gamma models. *Stat. Methods Appl.* **27** 69–89. MR3769304

[8] Benites, L., Maehara, R., Vilca, F. and Marmolejo-Ramos, F. (2017). Finite mixture of Birnbaum-Saunders distributions using the k-bumps algorithm. *arXiv preprint* arXiv:1708.00476.

[9] Birnbaum, Z. W. and Saunders, S. C. (1969). A new family of life distributions. *J. Appl. Probab.* **6** 319–327. MR0253493

[10] Böhning, D., Dietz, E., Schaub, R., Schlattmann, P. and Lindsay, B. G. (1994). The distribution of the likelihood ratio for mixtures of densities from the one-parameter exponential family. *Ann. Inst. Stat. Math.* **46** 373–388.

[11] Bourguignon, M., Silva, R. B. and Cordeiro, G. M. (2014). A new class of fatigue life distributions. *J. Stat. Comput. Simul.* **84** 2619–2635. MR3250961

[12] Cordeiro, G. M. and Lemonte, A. J. (2011). The β-Birnbaum-Saunders distribution: an improved distribution for fatigue life modeling. *Comput. Stat. Data Anal.* **55** 1445–1461. MR2741426

[13] Cordeiro, G. M., Lemonte, A. J. and Ortega, E. M. (2013). An extended fatigue life distribution. *Statistics* **47** 626–653. MR3060936

[14] Cordeiro, G. M. and Lemonte, A. J. (2014). The exponentiated generalized Birnbaum-Saunders distribution. *Appl. Math. Comput.* **247** 762–779. MR3270881

[15] Cordeiro, G. M., Lima, C. S., Cysneiros, A. H., Pascoa, M. A., Pescim, R. R. and Ortega, E. M. (2016). An extended Birnbaum-Saunders distribution: Theory, estimation, and applications. *Comm. Statist. Theo. and Meth.* **45** 2268–2297. MR3480648

[16] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. B* 1–38. MR0501537

[17] Díaz-García, J. A. and Leiva-Sánchez, V. (2005). A new family of life distributions based on the elliptically contoured distributions. *J. Stat. Plan. Infer.* **128** 445–457. MR2102769

[18] Hashemi, F., Amirzadeh, V. and Jamalizadeh, A. (2015). An Extension of the Birnbaum-Saunders distribution based on skew-normal t distribution. *JSRI.* **12** 1–37.

[19] Jamalizadeh, A., Hashemi, F. and Naderi, M. (2019). Discussion of "Birnbaum-Saunders distribution: A review of models, analysis, and applications". *Appl. Stochastic Models Bus. Ind.* **35** 82–89.

[20] Kappenman, R. F. (1975). Conditional confidence intervals for double exponential distribution parameters. *Technometrics* **17** 233–235. MR0370902

[21] Kenney, J. F. and Keeping, E. S. (1962). *Math. of Stat. Part 1*, third ed., Princeton, NJ. MR0137185

[22] Khosravi, M., Leiva, V., Jamalizadeh, A. and Porcu, E. (2016). On a nonlinear Birnbaum-Saunders model based on a bivariate construction and its characteristics. *Comm. Statist. Theo and Meth.* **45** 772–793. MR3453012

[23] Kundu, D., Kannan, N. and Balakrishnan, N. (2008). On the hazard function of Birnbaum-Saunders distribution and associated inference. *Comput. Stat. Data Anal.* **52** 2692–2702. MR2419535

[24] Leiva, V., Marchant, C., Saulo, H., Aslam, M. and Rojas, F. (2014). Capability indices for Birnbaum-Saunders processes applied to electronic and food industries. *J. Appl. Stat.* **41** 1881–1902. MR3292616

[25] Leiva, V., Vilca, F., Balakrishnan, N. and Sanhueza, A. (2010). A skewed sinh-normal distribution and its properties and application to air pollution. *Comm. Stat. Theo. M.* **39** 426–443. MR2745286

[26] Lemonte, A. J., Martínez-Florez, G. and Moreno-Arenas, G. (2015). Multivariate Birnbaum-Saunders distribution: Properties and associated inference. *J. Appl. Stat.* **85** 374–392. MR3270682

[27] Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. *J. R. Stat. Soc. Ser. B* **20** 102–107. MR0095550

[28] Lindsay, B. G. (1955). *Mixture models: theory, geometry and applications.* Institute of Mathematical Statistics, Hayward.

[29] Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *J. Roy. Stat. Soc. B* 226–233. MR0676213

[30] Martínez-Flórez, G., Bolfarine, H. and Gómez, H. W. (2014). An alpha-power extension for the Birnbaum-Saunders distribution. *Stat.* **48** 896–912. MR3234069

[31] Marchant, C., Leiva, V., Cysneiros, F. J. A. and Vivanco, J. F. (2016). Diagnostics in multivariate generalized Birnbaum-Saunders regression models. *J. Appl. Stat.* **43** 1–21. MR3546117

[32] Marchant, C., Leiva, V., Cysneiros, F. J. A. and Liu, S. (2018). Robust multivariate control charts based on Birnbaum-Saunders distributions. *J. Stat. Comput. Simul.* **88** 182–202. MR3737474

[33] McNeil, A. J., Frey, R. and Embrechts, P. (2005). *Quantitative risk management: Concepts, Techniques and Tools.* Princeton University Press. MR2175089

[34] McNicholas, P. D., Murphy, T. B., McDaid, A. F. and Frost, D. (2010). Serial and parallel implementations of model-based clustering via parsimonious Gaussian mixture models. *Comput. Stat. Data Anal.* **54** 711–723. MR2744427

[35] Meilijson, I. (1989). A fast improvement to the EM algorithm on its own terms. *J. Roy. Stat. Soc. B* 127–138. MR0984999

[36] Meng, X. L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **80** 267–278. MR1243503
[37] Moors, J. J. A. (1998). A quantile alternative for kurtosis. *J. R. Stat. Soc. Ser.* **37** 25–32.
[38] Naderi, M., Arabpour, A. and Jamalizadeh, A. (2018). Multivariate normal mean-variance mixture distribution based on Lindley distribution. *Comm. Statist. Simulation Comput.* **48** 1179–1192. MR3812405
[39] Ng, H. K. T., Kundu, D. and Balakrishnan, N. (2003). Modified moment estimation for the two parameter Birnbaum-Saunders distribution. *Comput. Statist. Data Anal.* **43** 283–298. MR1996813
[40] Raftery, A. E. (1995). Bayesian model seletion in social research. *Sociol. Methods* **25** 111–163.
[41] Olmos, N. M., Martínez-Flórez, G. and Bolfarine, H. (2017). Bimodal Birnbaum-Saunders distribution with applications to non negative measurements. *Comm. Statist. Theo Meth.* **46** 6240–6257. MR3631510
[42] Ortega, E. M., Lemonte, A. J., Cordeiro, G. M. and Nilton da Cruz, J. (2016). The odd Birnbaum-Saunders regression model with applications to lifetime data. *J. Stat. Theo Pract.* **10** 780–804. MR3558402
[43] Punzo, A., Mazza, A. and Maruotti, A. (2018). Fitting insurance and economic data with outliers: a flexible approach based on finite mixtures of contaminated gamma distributions. *J. Appl. Stat.* **45** 2563–2584. MR3851143
[44] Vilca-Labra, F. and Leiva-Sánchez, V. (2006). A new fatigue life model based on the family of skew-elliptical distributions. *Comm. Statist. Theo Meth* **35** 229–244. MR2274046
[45] Vilca, F., Santana, L., Leiva, V. and Balakrishnan, N. (2011). Estimation of extreme percentiles in Birnbaum-Saunders distributions. *Comput. Stat. Data Anal.* **55** 1665–1678. MR2748670

Farzane Hashemi
Department of Statistics
Faculty of Mathematics and Computer
Shahid Bahonar University of Kerman
Kerman
Iran
E-mail address: farzane.hashemi1367@yahoo.com

Mehrdad Naderi
Department of Statistics
Faculty of Natural & Agricultural Sciences
University of Pretoria
Pretoria
South Africa
E-mail address: m.naderi@up.ac.za

Ahad Jamalizadeh
Department of Statistics
Faculty of Mathematics and Computer
Shahid Bahonar University of Kerman
Kerman
Iran
E-mail address: a.jamalizadeh@uk.ac.ir