



On the Extended Birnbaum–Saunders Distribution Based on the Skew-*t*-Normal Distribution

Tahereh Poursadeghfard¹ · Ahad Jamalizadeh² · Alireza Nematollahi¹

Received: 26 July 2017 / Accepted: 18 July 2018 / Published online: 31 July 2018
© Shiraz University 2018

Abstract

In this article, a generalized version of the univariate Birnbaum–Saunders distribution based on the skew-*t*-normal distribution is introduced and its characterizations, properties are studied. Maximum likelihood estimation of the parameters via the ECM algorithm evaluated by Monte Carlo simulations is also discussed. Finally, two real datasets are analyzed for illustrative purposes.

Keywords Birnbaum–Saunders distribution · ECM algorithm · Information matrix · Skew-*t*-normal distribution

1 Introduction

The two-parameter Birnbaum–Saunders (BS) distribution as a life distribution was originally introduced by Birnbaum and Saunders (1969) as a failure model due to cracks. A random variable U is said to have the BS distribution with shape and scale parameters $\alpha > 0, \beta > 0$, respectively, if its cumulative distribution function (cdf) and probability density function (pdf) are given by

$$F_{BS}(u; \alpha, \beta) = \Phi(a(u; \alpha, \beta)), \quad u > 0,$$
$$f_{BS}(u; \alpha, \beta) = \phi(a(u; \alpha, \beta))A(u; \alpha, \beta), \quad u > 0,$$

respectively, where $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cdf and pdf of the standard normal distribution, respectively, and

$$a(u; \alpha, \beta) = \frac{1}{\alpha} \left(\sqrt{\frac{u}{\beta}} - \sqrt{\frac{\beta}{u}} \right) \quad \text{and} \quad A(u; \alpha, \beta) = \frac{da(u; \alpha, \beta)}{du} =$$

$\frac{u+\beta}{2\alpha\sqrt{\beta}\sqrt{u^3}}$. The stochastic representation of U is given by

$$U \stackrel{d}{=} \frac{\beta}{4} \left[\alpha Z + \sqrt{(\alpha Z)^2 + 4} \right]^2, \quad (1)$$

where $\stackrel{d}{=}$ means equal in distribution, $Z \sim N(0, 1)$ and consequently $Z \stackrel{d}{=} \frac{1}{\alpha} \left(\sqrt{\frac{u}{\beta}} - \sqrt{\frac{\beta}{u}} \right)$.

The BS distribution as a skew distribution has been frequently applied in the last few years to biological model by Desmond (1985), to the medical field by Leiva et al. (2007) and Barros et al. (2008) and to the forestry and environmental sciences by Podaski (2008), Leiva et al. (2010) and Vilca et al. (2011).

For more flexibility, several extensions of the BS distribution have been considered in the literature. For example, one can refer to Diaz-Garcia and Leiva-Sanchez (2005), Sanhueza et al. (2008), Leiva et al. (2008) and Gomez et al. (2009).

The well-known skew-normal (SN) distribution introduced by Azzalini (1985, 1986) could be used instead of the usual normal distribution, whenever the data present skewness. In this case, percentiles concentrated on the left-tail or right-tail of the distribution should be predicted in a better way.

A random variable Y is said to have the standard SN distribution with shape parameter $\lambda \in R$, denoted by $Y \sim SN(\lambda)$, if its pdf is given by

✉ Alireza Nematollahi
ar.nematollahi@shirazu.ac.ir

Tahereh Poursadeghfard
t.poursadeghfard@gmail.com

Ahad Jamalizadeh
a.jamalizadeh@uk.ac.ir

¹ Department of Statistics, College of Sciences, Shiraz University, Shiraz, Iran

² Department of Statistics, Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Kerman, Iran

$$f_{SN}(y; \lambda) = 2\phi(y)\Phi(\lambda y), \quad y \in R.$$

Vilca et al. (2011) considered the SN(λ) for the random variable Z in (1) and obtained the skew-normal Birnbaum–Saunders (SN-BS) distribution with the pdf

$$\begin{aligned} f_{SN-BS}(u; \alpha, \beta, \lambda) &= f_{SN}(a(u; \alpha, \beta); \lambda)A(u; \alpha, \beta) \\ &= 2\phi(a(u; \alpha, \beta))\Phi(\lambda a(u; \alpha, \beta))A(u; \alpha, \beta), \\ u, \alpha, \beta &> 0, \lambda \in R. \end{aligned}$$

The maximum likelihood estimations of the SN-BS distribution parameters are usually obtained by ECM algorithm. Vilca et al. (2011) have shown that the extreme percentiles can be predicted with high accuracy by using their proposed model.

Also Hashmi et al. (2015) considered SNT distribution [see Nadarajah and Kotz (2003)] for the random variable Z in (1) and obtained some better results. The proposed pdf is

$$\begin{aligned} f_{SNT-BS}(u; \alpha, \beta, \lambda, v) &= 2\phi(a(u; \alpha, \beta)) \\ T(\lambda a(u; \alpha, \beta); v)A(u; \alpha, \beta), \quad u, \alpha, \beta, v > 0, \lambda \in R, \end{aligned}$$

where $T(·; v)$ denotes the cdf of the Student’s t -distribution.

Gomez et al. (2007) introduced the class of distributions, called skew symmetric distribution including the skew- t -normal (STN) distribution and showed that it fits well to model data with heavy tail and strong asymmetry. A Bayesian approach to the study of the scale mixtures log-Birnbaum–Saunders regression models with censored data is also proposed by Lachos et al. (2017).

In this paper, we extend the BS distribution based on the skew- t -normal distribution, called skew- t -normal Birnbaum–Saunders distribution (STN-BS), and show that extreme percentiles can be better predicted rather than some other extensions of the Birnbaum Saunders distribution.

The rest of this paper is organized as follows. Section 2 defines a new version of the BS distribution and presents a useful stochastic representation, where several properties for the proposed distribution are also given. Section 3 concerns with the estimation of the parameters by maximum likelihood method via the ECM algorithm, where the Fisher information matrix is also calculated. Finally in Sect. 4, we illustrate the proposed methodology by analyzing two real datasets.

2 The STN-BS model and some characterizations

In this section, we consider the STN distribution to define the BS distribution based on STN distribution and derive some of its properties. Following Gomez et al. (2007), recall that a random variable Y is said to have the STN distribution with skewness parameter $\lambda \in R$ and degree of freedom $v \in (0, \infty)$, denoted by $Y \sim \text{STN}(\lambda, v)$, if its pdf

$$f_{STN}(y; \lambda, v) = 2t(y; v)\Phi(\lambda y), \quad y \in R, \tag{2}$$

where $t(y; v)$ denotes the pdf of the Student’s t -distribution with degree of freedom v . The density (2) reduces to the $t(y; v)$ distribution when $\lambda = 0$, to the truncated Student’s t -distribution, when $|\lambda| \rightarrow \infty$, and to the SN distribution, when $v \rightarrow \infty$. Now, a BS distribution based on STN distribution is easily defined, as given in the following definition.

Definition 2.1 A random variable U is said to have the STN-BS distribution with parameter $(\alpha, \beta, \lambda, v)$ and is denoted by $U \sim \text{STN-BS}(\alpha, \beta, \lambda, v)$, if it has the following stochastic representation

$$U \stackrel{d}{=} \frac{\beta}{4} \left[\alpha Y + \sqrt{(\alpha Y)^2 + 4} \right]^2,$$

where $Y \sim \text{STN}(\lambda, v)$. Then, the pdf of U can be easily

obtained as

$$\begin{aligned} f_{STN-BS}(u; \alpha, \beta, \lambda, v) &= f_{STN}(a(u; \alpha, \beta); \lambda, v)A(u; \alpha, \beta) \\ &= 2t(a(u; \alpha, \beta); v)\Phi(\lambda a(u; \alpha, \beta))A(u; \alpha, \beta), \\ u, \alpha, \beta, v > 0, \lambda \in R. \end{aligned} \tag{3}$$

2.1 Simple properties and moments

In this section, we present some simple properties and expressions for the moments of STN-BS distributions.

1. For $\lambda = 0$, the pdf in (3) reduces to the pdf of the T-BS distribution which is an extension of BS distribution obtained by replacing the random variable Z in (1) with the Student’s t random variable with degree of freedom v . In this case, the pdf is given by

$$f_{T-BS}(u; v) = t(a(u; \alpha, \beta); v)A(u; \alpha, \beta).$$

2. The pdf in (3) tends to the pdf of the SN-BS distribution, as $v \rightarrow +\infty$

3. If $U \sim \text{STN-BS}(\alpha, \beta, \lambda, v)$, then $U^{-1} \sim \text{STN-BS}(\alpha, \beta^{-1}, -\lambda, v)$ and $cU \sim \text{STN-BS}(\alpha, c\beta, \lambda, v)$, for $c > 0$.
4. If $U \sim \text{STN-BS}(\alpha, \beta, \lambda, v)$, then $V \stackrel{d}{=} \left| \frac{1}{\alpha} \left(\sqrt{\frac{U}{\beta}} - \sqrt{\frac{\beta}{U}} \right) \right| \sim \text{HT}(v)$, where $\text{HT}(v)$ denotes the Student's half- t -distribution with degree of freedom v .
5. If $U_{T^*} \sim \text{T-BS}(\alpha, \beta, v)$ and $T^* \sim t(., v)$, then the mean, variance, coefficient of variation (CV), skewness (CS) and kurtosis (CK) of U_{T^*} denoted by $E[U_{T^*}]$, $V[U_{T^*}]$, $\gamma[U_{T^*}]$, $\alpha_3[U_{T^*}]$ and $\alpha_4[U_{T^*}]$ are given by

$$\begin{aligned}
 E[U_{T^*}] &= \frac{\beta}{2} \left[\alpha^2 ET^{*2} + 2 \right], \\
 V[U_{T^*}] &= \frac{\beta^2 \alpha^2}{4} \left[\alpha^2 (2ET^{*4} - E^2 T^{*2}) + 4ET^{*2} \right], \\
 \gamma[U_{T^*}] &= \frac{\alpha \sqrt{\alpha^2 (2ET^{*4} - E^2 T^{*2}) + 4ET^{*2}}}{[\alpha^2 ET^{*2} + 2]}, \\
 \alpha_3[U_{T^*}] &= \frac{1}{[V[U_{T^*}]]^{\frac{3}{2}}} \frac{\beta^3 \alpha^4}{8} \\
 &\quad \left[\alpha^2 (4ET^{*6} - 6ET^{*4} ET^{*2} + 2E^3 T^{*2}) \right. \\
 &\quad \left. + 12ET^{*4} - 12E^2 T^{*2} \right], \\
 \alpha_4[U_{T^*}] &= \frac{1}{[V[U_{T^*}]]^2} \frac{\beta^4 \alpha^4}{8} \\
 &\quad \left[\alpha^4 (8ET^{*8} - 16ET^{*2} ET^{*6} + 12ET^{*4} E^2 T^{*2} - 3E^4 T^{*2}) \right. \\
 &\quad \left. + \alpha^2 (32ET^{*6} - 48ET^{*4} ET^{*2} + 24E^3 T^{*3}) + 16ET^{*4} \right].
 \end{aligned}$$

respectively, where

$$\begin{aligned}
 ET^{*2} &= \frac{v}{v-2}, \quad v > 2, \\
 ET^{*4} &= \frac{3v^2}{(v-2)(v-4)}, \quad v > 4, \\
 ET^{*6} &= \frac{15v^3}{(v-2)(v-4)(v-6)}, \quad v > 6, \\
 ET^{*8} &= \frac{105v^4}{(v-2)(v-4)(v-6)(v-8)}, \quad v > 8.
 \end{aligned}$$

The moments of the STN-BS distribution can be expressed in terms of the moments of T-BS distribution. In the following proposition, we present the relationships between the means, variances, coefficients of variation, skewness and kurtosis of the STN-BS and T-BS distributions.

Proposition 2.1 *Let $U \sim \text{STN-BS}(\alpha, \beta, \lambda, v)$ and $U_{T^*} \sim \text{T-BS}(\alpha, \beta, v)$. Then, the mean, variance, coefficient of variation, coefficient of skewness and the coefficient of kurtosis denoted by $E[U]$, $V[U]$, $\gamma[U]$, $\alpha_3[U]$, and $\alpha_4[U]$, of U in terms of U_{T^*} are given by*

$$\begin{aligned}
 E[U] &= E[U_{T^*}] + \frac{\alpha\beta}{2} \omega_1, \\
 V[U] &= V[U_{T^*}] + \left[\frac{\alpha\beta}{2} \right]^2 \alpha_\omega, \\
 \gamma[U] &= \gamma[U_{T^*}] \frac{\sqrt{1 + \frac{\alpha^2 \beta^2 \alpha_\omega}{V(2U_{T^*})}}}{1 + \frac{\alpha\beta\omega_1}{E[2U_{T^*}]}}}, \\
 \alpha_3[U] &= \alpha_3[U_{T^*}] \left[\frac{\alpha^2 (2ET^{*4} - E^2 T^{*2}) + 4ET^{*2}}{\alpha^2 (2ET^{*4} - E^2 T^{*2}) + 4ET^{*2} + \alpha\omega} \right]^{\frac{3}{2}} \\
 &\quad + \frac{2[a_0 + a_1\alpha + a_2\alpha^2]}{[\alpha^2 (2ET^{*4} - E^2 T^{*2}) + 4ET^{*2} + \alpha\omega]^{\frac{3}{2}}}, \\
 \alpha_4[U] &= \left[\alpha_4[U_{T^*}] + \frac{b_0 + b_1\alpha + b_2\alpha^2 + b_3\alpha^3}{(\alpha^2 (2ET^{*4} - E^2 T^{*2}) + 4ET^{*2})^2} \right] \\
 &\quad \frac{[\alpha^2 (2ET^{*4} - E^2 T^{*2}) + 4ET^{*2}]^2}{[\alpha^2 (2ET^{*4} - E^2 T^{*2}) + 4ET^{*2} + \alpha\omega]^2},
 \end{aligned}$$

respectively, where

$$\begin{aligned}
 \alpha_\omega &= 2\alpha(\omega_3 - \omega_1 ET^{*2}) - \omega_1^2, \\
 a_0 &= \omega_1^3 - 6\omega_1 ET^{*2} + 2\omega_3, \\
 a_1 &= 3\omega_1^2 ET^{*2} - 3\omega_1 \omega_3, \\
 a_2 &= 2\omega_5 - 3\omega_3 ET^{*2} - 3\omega_1 ET^{*4} + 3\omega_1 E^2 T^{*2}, \\
 b_0 &= -3\omega_1^4 - 16\omega_1 \omega_3 + 24\omega_1^2 ET^{*2}, \\
 b_1 &= 16\omega_5 - 16\omega_3 ET^{*2} - 48\omega_1 ET^{*2} + 48\omega_1 E^2 T^{*2} \\
 &\quad + 12\omega_1^2 \omega_3 - 12\omega_1^3 ET^{*2}, \\
 b_2 &= -16\omega_1 \omega_5 + 12\omega_1^2 ET^{*4} - 18\omega_1^2 E^2 T^{*2} + 24\omega_1 \omega_3 ET^{*2}, \\
 b_3 &= 8\omega_7 - 16\omega_1 ET^{*6} - 16\omega_5 ET^{*2} + 12\omega_3 E^2 T^{*2} \\
 &\quad + 24\omega_1 ET^{*2} ET^{*4} - 12\omega_1 E^3 T^{*2}, \\
 \omega_k &= E[Y^k \sqrt{\alpha^2 Y^2 + 4}], \quad k = 1, 3, 5, 7,
 \end{aligned}$$

and $Y \sim \text{STN}(\lambda, v)$, $T^* \sim t(v)$. For calculating the values for ω_k , the involved integrals must be solved by using some numerical methods. We have applied the integrate function in the statistical software R.

Table 1 provides values for the mean (μ), standard deviation (SD), CS and CK of the $\text{STN-BS}(\alpha, 1, \lambda, 9)$ for different values of α and λ .

We observe that, for both positive values of λ and large values of α , the distribution has very large kurtosis.

Figure 1 displays the graph of the densities, BS, SN-BS, SNT-BS, T-BS and STN-BS, for some of the selected values of parameters.

2.2 Some useful results

Here, we provide some useful results which will be used in the estimation methods. Following Cabral et al. (2008) and

Table 1 The values for the mean (μ), standard deviation (SD), CS and CK of the STN-BS($\alpha, 1, \lambda, 9$)

$ \lambda $	α	μ		SD		CS		CK	
		λ	$-\lambda$	λ	$-\lambda$	λ	$-\lambda$	λ	$-\lambda$
.05	.05	1.0042	.9990	.0570	.0566	.3287	.2166	4.5534	4.3089
	.10	1.0116	1.0013	.1442	.1137	.6168	.4907	5.5913	4.8840
	.50	1.1900	1.1314	.7406	.6891	3.5077	3.0403	76.7088	36.1312
	1.00	1.7181	1.5676	2.1895	1.9704	6.2129	5.5251	223.1611	103.2840
	1.50	2.5903	2.3025	4.5283	4.0253	7.4089	6.6907	304.7121	143.6374
.20	.05	1.0115	.9917	.0568	.05513	.4603	.0620	4.71245	4.0006
	.10	1.0264	.9865	.1164	.1095	.7471	.31107	5.9762	4.1919
	.50	1.2738	1.0477	.7945	.6069	3.5971	2.3862	70.8695	17.6882
	1.00	1.9316	1.3541	2.4222	1.6291	6.011	4.5057	181.8941	49.8618
	1.50	2.9967	1.8962	5.0622	3.2431	7.0136	5.6109	238.0859	72.1601
.50	.05	1.0233	.9800	.0542	.05092	.6095	-.1424	4.8306	3.7370
	.10	1.0499	.9630	.1126	.0994	.8729	.0711	6.2268	3.6334
	.50	1.4036	.9178	.8260	.4742	3.3987	1.6690	61.9214	8.8838
	1.00	2.2523	1.0334	2.5901	1.1118	5.3482	3.4545	142.6597	25.5825
	1.5	3.5949	1.2980	5.4609	2.0713	6.1233	4.6155	180.5355	41.3649
1.00	.05	1.0340	.9692	.0488	.04452	.8230	-.3885	5.4579	3.8701
	.10	1.0715	.9414	.1028	.0855	1.0759	-.1977	7.2391	3.5408
	.50	1.5187	.8027	.8046	.3486	3.4861	1.0900	66.9406	5.2713
	1.00	2.5202	.7656	2.5862	.6783	5.2311	2.5871	141.7391	14.6409
	1.5	4.0730	.8198	5.4960	1.1191	5.9001	3.8660	174.3631	28.0149
1.50	.05	1.0389	.9644	.04434	.0511	1.0745	-.6147	6.2723	4.2286
	.10	1.0812	.9317	.0959	.0771	1.2998	-.4306	8.4625	3.7360
	.50	1.5690	.7524	.7804	.2876	3.7123	.7254	74.5579	3.9733
	1.00	2.6308	.6549	2.5481	.4929	5.3750	2.0036	149.4574	9.7846
	1.50	4.2607	.6321	5.4436	.7306	5.9919	3.2492	180.3405	20.4836
2.00	.05	1.0413	.9620	.0431	.0383	1.2251	-.7977	6.9558	4.5740
	.10	1.0860	.9269	.0919	.0724	1.4817	-.6175	9.4543	3.9654
	.50	1.5933	.7280	.7651	.2554	3.8921	.4591	80.1300	3.3315
	1.00	2.6828	.6029	2.5224	.4028	5.4980	1.5691	155.1552	7.0311
	1.50	4.3457	.5471	5.4067	.5498	6.0795	2.6918	184.9133	14.9291

Ho et al. (2011), the following convenient stochastic representation holds for $Y \sim \text{STN}(\lambda, v)$,

$$Y^d = \left[\frac{\lambda |Z_1|}{\sqrt{\tau(\tau + \lambda^2)}} + \frac{Z_2}{\sqrt{\tau + \lambda^2}} \right], \tag{4}$$

where Z_1, Z_2 are two independent $N(0, 1)$ and $\tau \sim \Gamma(\frac{v}{2}, \frac{v}{2})$

(the gamma distribution with shape parameter $\frac{v}{2}$ and scale parameter $\frac{v}{2}$) is independent of Z_1 and Z_2 . Set $\gamma = \sqrt{\frac{\tau + \lambda^2}{\tau}} |Z_1|$, then (4) becomes

$$Y^d = \left[\frac{\lambda \gamma}{(\tau + \lambda^2)} + \frac{Z_2}{\sqrt{\tau + \lambda^2}} \right]. \tag{5}$$

So the following representation for STN-BS random variable U holds,

$$U^d = \frac{\beta}{4} \left[\alpha \left\{ \frac{\lambda \gamma}{(\tau + \lambda^2)} + \frac{Z_2}{\sqrt{\tau + \lambda^2}} \right\} + \sqrt{\left\{ \alpha \left(\frac{\lambda \gamma}{(\tau + \lambda^2)} + \frac{Z_2}{\sqrt{\tau + \lambda^2}} \right) \right\}^2 + 4} \right]^2.$$

The following two propositions are useful for the ML estimation of the STN-BS distribution parameters via ECM algorithm discussed in the next section. The proofs are collected in “Appendix.”

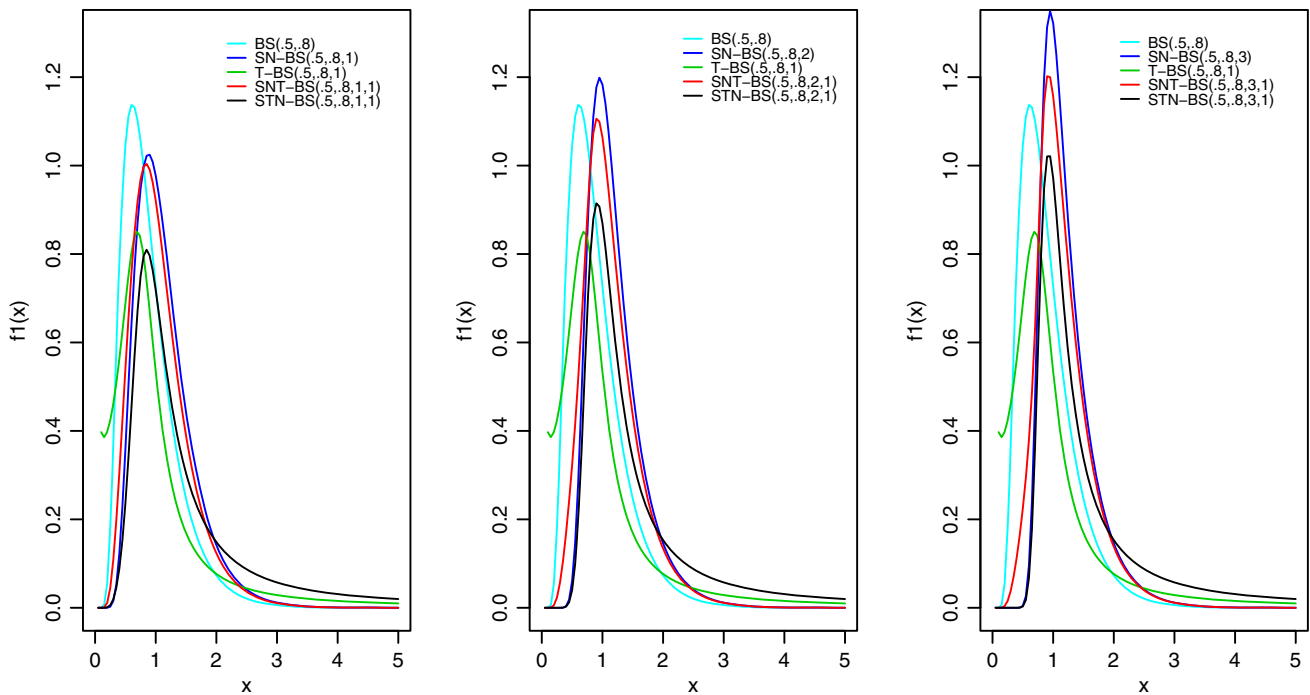


Fig. 1 The graph of the densities; BS, SN-BS, T-BS, SNT-BS and STN-BS for selected values of parameters

Proposition 2.2 Let $U \sim \text{STN-BS}$ and $\gamma = \sqrt{\frac{\tau + \lambda^2}{\tau}}|Z_1|$ and $\tau \sim \Gamma(\frac{v}{2}, \frac{v}{2})$, then the distributions of $U | (\gamma, \tau)$ and $\gamma | \tau$ are given by

$$U | (\gamma, \tau) \sim \text{EBS}\left(\frac{\alpha}{\sqrt{\tau + \lambda^2}}, \beta, 2, \frac{-\lambda\gamma}{\sqrt{\tau + \lambda^2}}\right),$$

and

$$\gamma | \tau \sim \text{TN}\left(0, \frac{\tau + \lambda^2}{\tau}, (0, +\infty)\right),$$

respectively, where $\text{EBS}(\alpha, \beta, \sigma, \lambda)$ denotes the extended BS distribution discussed by Leiva et al. (2010) and $\text{TN}(\mu, \sigma^2; (a, b))$ denotes the truncated normal distribution for $N(\mu, \sigma^2)$ lying within the truncated interval (a, b) .

Proposition 2.3 (a) The conditional expectation of τ given $U = u$ is

$$E(\tau | U = u) = \frac{v + 1}{v + a^2(u; \alpha, \beta)}.$$

(b) The conditional expectation of $\log \tau$ given $U = u$ is

$$E(\log \tau | U = u) = DG\left(\frac{v + 1}{2}\right) - \log\left(\frac{v + a^2(u; \alpha, \beta)}{2}\right),$$

where $DG(x) = \frac{d}{dx} \log \Gamma(x)$ is the digamma function.

(c) The conditional expectation of γ given $U = u$ is

$$E(\gamma | U = u) = \lambda a(u; \alpha, \beta) + \frac{\phi(\lambda a(u; \alpha, \beta))}{\Phi(\lambda a(u; \alpha, \beta))}.$$

3 Maximum likelihood estimation

The EM-based algorithms are a multi-step optimization method to build a sequence of easier maximization problems whose limit is the answer to the original problem. Each iteration of the EM-algorithm contains two steps: the Expectation step or the E-step and the Maximization step or the M-step. The literature on EM-based algorithms and their applications is quite rich. For a comprehensive listing of the important references, details and more information, we refer the readers to Dempster et al. (1977), Meng and Rubin (1993) and McLachlan and Krishnan (2008) and references therein. In this part, we derive the ML estimation of the STN-BS distribution parameters via modification of the EM-algorithm [Expectation/Conditional Maximization (ECM) algorithm].

3.1 Estimation via ECM algorithm

Let $\mathbf{U} = [U_1, \dots, U_n]^T$ be a random sample of size n from $\text{STN-BS}(\alpha, \beta, \lambda, v)$. From Proposition 2.2, we set the observed data by $\mathbf{u} = [u_1, \dots, u_n]^T$, the missing data by $\boldsymbol{\tau} = [\tau_1, \dots, \tau_n]^T$ and $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_n]^T$, and the complete data by $\mathbf{u}^{(c)} = [\mathbf{u}^T, \boldsymbol{\tau}^T, \boldsymbol{\gamma}^T]^T$.

Then, we construct the complete data log-likelihood function of $\theta = (\alpha, \beta, \lambda, v)$ given $\mathbf{u}^{(c)}$, ignoring additive constant terms, as follows:

$$\begin{aligned} \ell^{(c)}(\theta|\mathbf{u}^{(c)}) &= \sum_{i=1}^n \ell_i^{(c)}(\theta|(u_i, \tau_i, \gamma_i)) \\ &= \frac{\lambda}{\alpha} \sum_{i=1}^n \gamma_i \varepsilon(u_i; \beta) - \frac{1}{2} \frac{\lambda^2}{\alpha^2} \sum_{i=1}^n \varepsilon^2(u_i; \beta) \\ &\quad - \frac{1}{2\alpha^2} \sum_{i=1}^n \tau_i \varepsilon^2(u_i; \beta) - \frac{v}{2} \sum_{i=1}^n \tau_i \\ &\quad + \sum_{i=1}^n \log \frac{u_i + \beta}{\sqrt{\beta}} - n \log(\alpha) \\ &\quad + \frac{nv}{2} \log\left(\frac{v}{2}\right) - n \log \Gamma\left(\frac{v}{2}\right) + \frac{v-1}{2} \sum_{i=1}^n \log \tau_i. \end{aligned}$$

Suppose $\hat{\theta}^{(r)} = (\hat{\alpha}^{(r)}, \hat{\beta}^{(r)}, \hat{\lambda}^{(r)}, \hat{v}^{(r)})$ is the current estimate (in the r th iteration) of θ . Based on the ECM algorithm principle, in the E-step, we should first form the following conditional expectation

$$\begin{aligned} \mathbf{Q}(\theta|\hat{\theta}^{(r)}) &= E(\ell^{(c)}(\theta|\mathbf{u}^{(c)})) \\ &= \frac{\lambda}{\alpha} \sum_{i=1}^n \hat{S}_{3i}^{(r)} \varepsilon(u_i; \beta) - \frac{1}{2} \frac{\lambda^2}{\alpha^2} \sum_{i=1}^n \varepsilon^2(u_i; \beta) \\ &\quad - \frac{1}{2\alpha^2} \sum_{i=1}^n \hat{S}_{1i}^{(r)} \varepsilon^2(u_i; \beta) - \frac{v}{2} \sum_{i=1}^n \hat{S}_{1i}^{(r)} \\ &\quad + \sum_{i=1}^n \log \frac{u_i + \beta}{\sqrt{\beta}} - n \log(\alpha) + \frac{nv}{2} \log\left(\frac{v}{2}\right) \\ &\quad - n \log \Gamma\left(\frac{v}{2}\right) + \frac{v-1}{2} \sum_{i=1}^n \hat{S}_{2i}^{(r)}, \end{aligned} \tag{6}$$

where

$$\hat{S}_{1i}^{(r)} = E(\tau_i | u_i, \hat{\theta}^{(r)}) = \frac{\hat{v}^{(r)} + 1}{\hat{v}^{(r)} + a^2(u_i; \hat{\alpha}^{(r)}, \hat{\beta}^{(r)})}, \tag{7}$$

$$\begin{aligned} \hat{S}_{2i}^{(r)} &= E(\log \tau_i | u_i, \hat{\theta}^{(r)}) = DG\left(\frac{\hat{v}^{(r)} + 1}{2}\right) \\ &\quad - \log \frac{\hat{v}^{(r)} + a^2(u_i; \hat{\alpha}^{(r)}, \hat{\beta}^{(r)})}{2}, \end{aligned} \tag{8}$$

$$\begin{aligned} \hat{S}_{3i}^{(r)} &= E(\gamma_i | u_i, \hat{\theta}^{(r)}) = \hat{\lambda}^{(r)} a(u_i; \hat{\alpha}^{(r)}, \hat{\beta}^{(r)}) \\ &\quad + \frac{\phi(\hat{\lambda}^{(r)} a(u_i; \hat{\alpha}^{(r)}, \hat{\beta}^{(r)})}{\Phi(\hat{\lambda}^{(r)} a(u_i; \hat{\alpha}^{(r)}, \hat{\beta}^{(r)})}. \end{aligned} \tag{9}$$

Then, the corresponding ECM algorithm is done as follows:

E-step Given $\theta = \hat{\theta}^{(r)}$, compute $\hat{S}_{1i}^{(r)}, \hat{S}_{2i}^{(r)}, \hat{S}_{3i}^{(r)}$, using Eqs. (7), (8) and (9) for $i = 1, \dots, n$.

CM-step1 Fix $\beta = \hat{\beta}^{(r)}$ and update $\hat{\alpha}^{(r)}, \hat{\lambda}^{(r)}$ by maximizing (6) over α and λ , which leads to

$$\begin{aligned} \hat{\lambda}^{(r+1)} &= \frac{\hat{\alpha}^{(r+1)} \sum_{i=1}^n \varepsilon(u_i; \hat{\beta}^{(r)}) \hat{S}_{3i}^{(r)}}{\sum \varepsilon^2(u_i; \beta)}, \\ \hat{\alpha}^{2(r+1)} &= \frac{1}{n} \sum_{i=1}^n \varepsilon^2(u_i; \hat{\beta}^{(r)}) \hat{S}_{1i}^{(r)}. \end{aligned}$$

CM-step 2 Fix $\alpha = \hat{\alpha}^{(r+1)}, \lambda = \hat{\lambda}^{(r+1)}, \beta = \hat{\beta}^{(r)}$ and update $\hat{v}^{(r)}$ by maximizing (6) over v , which leads to solve the root of the following equation

$$\frac{n}{2} \left(\log \frac{v}{2} + 1 - DG\left(\frac{v}{2}\right) + \frac{1}{n} \sum_{i=1}^n (\hat{S}_{2i}^{(r)} - \hat{S}_{1i}^{(r)}) \right) = 0.$$

CM-step 3 Fix $\alpha = \hat{\alpha}^{(r+1)}, \lambda = \hat{\lambda}^{(r+1)}, v = \hat{v}^{(r+1)}$ and update $\hat{\beta}^{(r)}$ using

$$\hat{\beta}^{(r+1)} = \arg \max Q(\hat{\alpha}^{(r+1)}, \beta, \hat{\lambda}^{(r+1)}, \hat{v}^{(r+1)} | \hat{\theta}^{(r)}).$$

Note that the CM-steps 2 and 3 require a one-dimensional search for the root of v and optimization with respect to β , which can be easily obtained by using the *uniroot* and the *optimize* functions in the R statistical language package version 3.3.1 (R Development Core Team 2016).

Remark 3.1 (i) In the representation (4), when $\tau = 1$, random variable T reduces to a random variable with SNBS distribution. See Vilca et al. (2011).

(ii) For ensuring that the true ML estimates are obtained, we recommend running the ECM algorithm using a range of different starting values and checking whether all of them result in the same estimates. Also, the initial estimates are obtained using numerical methods, such as procedure *DEoptim* in the statistical software R for maximizing the corresponding likelihood function.

3.2 The information matrix

Under some regularity conditions, the covariance matrix of ML estimates $\hat{\theta}$, can be approximated by the inverse of the observed information matrix, i.e., $I_0(\hat{\theta}|\mathbf{u}) = \frac{-\partial^2 \ell(\theta|\mathbf{u})}{\partial \theta \partial \theta^T} \Big|_{\theta=\hat{\theta}}$,

$$\begin{aligned} \ell(\theta|\mathbf{u}) &= \sum_{i=1}^n \ell_i(\theta|u_i) = n \log 2 + \sum_{i=1}^n \log t(a(u_i, \alpha, \beta); v) \\ &\quad + \sum_{i=1}^n \log \Phi(\lambda a(u_i, \alpha, \beta)) + \log(A(u_i, \alpha, \beta)). \end{aligned}$$

Now we use Basford et al. (1997) to obtain

$$I_0(\hat{\theta}|\mathbf{u}) = \sum_{i=1}^n \hat{\mathbf{S}}_i \hat{\mathbf{S}}_i^T,$$

where

$$\widehat{\mathbf{S}}_i = \frac{\partial \ell_i(\boldsymbol{\theta}|u_i)}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = (\widehat{\mathbf{S}}_{i,\alpha}, \widehat{\mathbf{S}}_{i,\beta}, \widehat{\mathbf{S}}_{i,\lambda}, \widehat{\mathbf{S}}_{i,v})^\top.$$

The elements of $\widehat{\mathbf{S}}_i$ are obtained by

$$\begin{aligned} \widehat{\mathbf{S}}_{i,\alpha} &= -\left(\frac{\widehat{v} + 1}{\widehat{v}}\right) \left(1 + \frac{a^2(u_i; \widehat{\alpha}, \widehat{\beta})}{\widehat{v}}\right)^{-1} a(u_i; \widehat{\alpha}, \widehat{\beta}) \frac{\partial}{\partial \alpha} a(u_i; \widehat{\alpha}, \widehat{\beta}) \\ &\quad + \frac{\widehat{\lambda} \phi(\widehat{\lambda} a(u_i; \widehat{\alpha}, \widehat{\beta}))}{\Phi(\widehat{\lambda}^{(r)} a(u_i; \widehat{\alpha}^{(r)}, \widehat{\beta}^{(r)}))} \frac{\partial}{\partial \alpha} a(u_i; \widehat{\alpha}, \widehat{\beta}) + \frac{\frac{\partial}{\partial \alpha} A(u_i, \widehat{\alpha}, \widehat{\beta})}{A(u_i, \widehat{\alpha}, \widehat{\beta})}, \\ \widehat{\mathbf{S}}_{i,\beta} &= -\left(\frac{\widehat{v} + 1}{\widehat{v}}\right) \left(1 + \frac{a^2(u_i; \widehat{\alpha}, \widehat{\beta})}{\widehat{v}}\right)^{-1} a(u_i; \widehat{\alpha}, \widehat{\beta}) \frac{\partial}{\partial \beta} a(u_i; \widehat{\alpha}, \widehat{\beta}) \\ &\quad + \frac{\widehat{\lambda} \phi(\widehat{\lambda} a(u_i; \widehat{\alpha}, \widehat{\beta}))}{\Phi(\widehat{\lambda} a(u_i; \widehat{\alpha}, \widehat{\beta}))} \frac{\partial}{\partial \beta} a(u_i; \widehat{\alpha}, \widehat{\beta}) + \frac{\frac{\partial}{\partial \beta} A(u_i, \widehat{\alpha}, \widehat{\beta})}{A(u_i, \widehat{\alpha}, \widehat{\beta})}, \\ \widehat{\mathbf{S}}_{i,\lambda} &= \frac{\phi(\widehat{\lambda} a(u_i; \widehat{\alpha}, \widehat{\beta}))}{\Phi(\widehat{\lambda} a(u_i; \widehat{\alpha}, \widehat{\beta}))} a(u_i; \widehat{\alpha}, \widehat{\beta}), \\ \widehat{\mathbf{S}}_{i,v} &= \frac{\frac{\partial}{\partial v} \tau(a(u_i; \widehat{\alpha}, \widehat{\beta}); \widehat{v})}{\tau(a(u_i; \widehat{\alpha}, \widehat{\beta}); \widehat{v})}. \end{aligned}$$

The covariance matrix can be useful for studying the asymptotic behavior of $\widehat{\boldsymbol{\theta}} = (\widehat{\alpha}, \widehat{\beta}, \widehat{\lambda}, \widehat{v})$ by the asymptotic normality of this ML estimator. Thus, we can form hypothesis tests and confidence regions for $\alpha, \beta, \lambda, v$ by using the multivariate normality of $\widehat{\boldsymbol{\theta}}$.

4 Simulation study and illustrative examples

4.1 Simulation study

We use simulations to evaluate the finite-sample performance of the ML estimates of STN-BS distribution parameters from the ECM algorithm described in Sect. 3. The sample sizes and true values of the parameters are $n = 50, 100$ and 500 , $\alpha = .1, .5, .75, 1.0$, $\lambda = .1, .2, .5$ and $v = 2$. The scale parameter is also taken as $\beta = 1$. In order to examine the performance of the ML estimates, for each sample size and for each estimate $\widehat{\theta}_i$, we compute the mean $E[\widehat{\theta}_i]$, the relative bias (RB) in absolute value given by $RB_i = |\{E[\widehat{\theta}_i] - \theta_i\} / \theta_i|$ and the root mean square error (\sqrt{MSE}) given by $\sqrt{E[\widehat{\theta}_i - \theta_i]^2}$, for $i = 1, 2, 3, 4$. The results for the ML estimates of the parameters α, β, λ and v are given in Tables 2, 3 and 4.

It is observed that, as we expect, the values of RB and RMSE of the ML estimators of the parameters decrease as the sample size increases.

4.2 Real data

In this section, two real datasets are applied in order to illustrate the applicability of the STN-BS model. For each dataset, we first obtain the ML estimations of the parameters via ECM algorithm described in Sect. 3. Then to compare competitions models, we use the maximized log-likelihood $\ell(\widehat{\boldsymbol{\theta}})$, and model selection criteria based on loss of information, such as the Akaike information criteria (AIC) and the Bayesian information criteria (BIC). Following Kass and Raftery (1995), we also use the Bayes factor (BF) to show more differences between the BIC values. Assuming that the data D have arisen from one of two hypothetical models, thus H_1 (model with a smaller BIC value) is contrasted to H_2 (model compared to H_1), according to $P(D | H_1)$ and $P(D | H_2)$. This factor can be obtained by using the following approximation proposed by Raftery (1995),

$$2 \log(B_{12}) \approx 2 \left[\log(P(D | \widehat{\theta}_1, H_1)) - \log(P(D | \widehat{\theta}_2, H_2)) \right] - [d_1 - d_2] \log(n),$$

where $P(D | \widehat{\theta}_1, H_1) = P(D | H_1)$, with $\widehat{\theta}_r$ being the ML estimate of θ_r under the model in H_r , d_r is the dimension of θ_r , for $r = 1, 2$, and n is the sample size; see Vilca et al. (2011). An interpretation of the BF is given in Table 5.

4.2.1 Ozone data

Table 6 presents a descriptive summary of Ozone data studied by Vilca et al. (2011), which are assumed to be uncorrelated and independent, including sample median, mean, standard deviation (SD), CV, CS and CK. As it is observed, the data come from a positively skewed distribution with a kurtosis greater than three. Thus, the STN-BS model can be suitable for these data.

Vilca et al. (2011) showed that the SN-BS distribution provides a better fit than the usual BS distribution. Now, we show that the proposed STN-BS distribution performs better than some other extensions of the Birnbaum Saunders distribution to fit on this dataset.

Estimation and model checking are provided in Table 7, which consists of the ML estimates and value of $\ell(\widehat{\boldsymbol{\theta}})$, AIC and BIC. Considering these values, we find that the STN-BS model provides a better fit than the other models. For these data, we also use the BF (approximated by the BIC) to contrast two following hypothesis testing:

- (i) $H_0^{(1)}$: STN-BS model versus $H_1^{(1)}$: BS model, which gives the value $2 \log B_{12}^{(1)} = 7.491$.

Table 2 Parameter estimates when $\lambda = .1$ and $\beta = 1.0, v = 2$

n	α	Mean	Estimates of α		Mean	Estimates of β		Mean	Estimates of λ		Mean	Estimates of v	
			RB	\sqrt{MSE}		RB	\sqrt{MSE}		RB	\sqrt{MSE}		RB	\sqrt{MSE}
50	.10	.1065	.0654	.0201	1.0065	.0064	.0365	.0617	.3831	.3170	2.5624	.2812	1.4308
	.50	.4627	.0746	.1180	1.0241	.0241	.2238	.4485	3.4845	.9566	1.8828	.0586	.6765
	.75	.8945	.1927	.3012	.8857	.1143	.3055	.5500	4.4999	.9271	2.9523	.4761	3.6270
	1.0	1.0588	.0588	.2082	1.0688	.0688	.4330	.3463	2.4630	.6621	2.5353	.2677	1.0915
100	.10	.1020	.0201	.0148	1.0013	.0013	.0206	.0919	.0806	.1211	2.3870	.1935	1.2867
	.50	.5167	.0333	.0938	.9782	.0218	.1313	.1709	.7085	.2902	2.2528	.1264	.9352
	.75	.7794	.0393	.1619	.9632	.0368	.2088	.2066	1.0656	.3969	2.1459	.0723	.6427
	1.0	1.0354	.0354	.1862	.9786	.0214	.2848	.1972	.9717	.3118	1.9400	.0300	.5141
500	.10	.1002	.0017	.0064	.9994	.0006	.0077	.1069	.0696	1.3642	2.0317	.0159	.7862
	.50	.5001	.0002	.0321	.9968	.0032	.0415	.1077	.0773	.0463	2.0230	.0115	.2429
	.75	.7505	.0006	.0477	.9937	.0063	.0662	.1091	.0911	.0025	2.0257	.0128	.2424
	1.0	1.0076	.0076	.0732	.9931	.0069	.0654	.1082	.0820	.0783	2.0217	.0109	.2015

Table 3 Parameter estimates when $\lambda = .2$ and $\beta = 1.0, v = 2$

n	α	Mean	Estimates of α		Mean	Estimates of β		Mean	Estimates of λ		Mean	Estimates of v	
			RB	\sqrt{MSE}		RB	\sqrt{MSE}		RB	\sqrt{MSE}		RB	\sqrt{MSE}
50	.10	.1447	.4469	.0447	.9356	.0644	.0644	.6210	2.1049	.4210	3.1932	.5966	1.1932
	.50	.5407	.0807	.1773	.9335	.0665	.1900	.5204	1.6025	.7950	2.7051	.3525	4.5663
	.75	.8715	.1620	.3949	.8961	.1039	.3738	.6840	2.4202	1.0886	2.4617	.2309	1.4109
	1.0	1.2547	.2547	.5899	.8427	.1573	.4997	.9495	3.7474	1.4972	1.9819	.0090	.8686
100	.10	.1133	.1330	.0138	.9826	.0174	.0188	.3260	.6295	.1320	2.6255	.3128	.6294
	.50	.5250	.0501	.1156	.9491	.0509	.1472	.3984	.9922	.5084	2.2230	.1115	.9186
	.75	.8394	.1192	.2026	.9203	.0797	.2465	.4794	1.3972	.7520	2.3515	.1757	.8740
	1.0	1.0907	.0907	.3253	.9271	.0729	.3460	.4917	1.4587	.7566	2.1862	.0931	.6390
500	.10	.1011	.0110	.0081	.9975	.0025	.0092	.2239	.1193	.0593	2.0380	.0190	.3012
	.50	.5022	.0043	.9186	.9952	.0048	.0482	.2137	.0685	.0725	2.0306	.0153	.4946
	.75	.7544	.0058	.0518	.9885	.0115	.0817	.2201	.10068	.0883	2.0241	.0121	.2405
	1.0	1.0089	.0089	.0791	.9850	.0150	.1253	.2270	.1351	.1295	2.0224	.0112	.2438

- (ii) $H_0^{(2)}$: STN-BS model versus $H_1^{(2)}$: SN-BS model which gives the value $2 \log B_{12}^{(2)} = 5.262$.
- (iii) $H_0^{(3)}$: STN-BS model versus $H_1^{(3)}$: SNT-BS model which gives the value $2 \log B_{12}^{(3)} = 1.1564$.
- (iv) $H_0^{(4)}$: STN-BS model versus $H_1^{(4)}$: T-BS model which gives the value $2 \log B_{12}^{(4)} = .838$.

According to Table 5, the above values of $2 \log B_{12}$ indicate “strong,” “positive,” “weak” and “weak” evidence in favor of $H_0^{(1)}, H_0^{(2)}, H_0^{(3)}$ and $H_0^{(4)}$.

The estimated density functions of the ozone data including the respected histograms, plotted in Fig. 2, and the PP plots and empirical and theoretical cdf plots given in Figs. 3, 4, 5, 6 and 7 confirm again the appropriateness of the STN-BS distribution.

4.2.2 Fatigue data

These data correspond to the fatigue life of 6061-T6 aluminum coupons cut parallel to the direction of rolling and oscillated at 18 cycles per second, were provided by Birnbaum and Saunders (1969). Table 8 presents a descriptive summary of fatigue data, which indicate that STN-BS model can be suitable in modeling this dataset.

We fit the STN-BS distribution to fatigue data and compare it to BS, SN-BS, SNT-BS and T-BS distributions. Estimation and model checking are provided in Table 9, which has the same configuration as given in Table 7. Considering the values of the $\ell(\hat{\theta})$, AIC and BIC given in this table, we find that the STN-BS model, once again provides a better fit than other models. Similarly to ozone data, we use the BF to highlight the differences between the values of the criteria presented in Table 9. We obtain

Table 4 Parameter estimates when $\lambda = .5$ and $\beta = 1.0, v = 2$

n	α	Mean	Estimates of α		Mean	Estimates of β		Mean	Estimates of λ		Mean	Estimates of v	
			RB	\sqrt{MSE}		RB	\sqrt{MSE}		RB	\sqrt{MSE}		RB	\sqrt{MSE}
50	.10	.0724	.2764	.0284	1.0320	.0320	.0340	.1313	.7374	.3707	2.4365	.2182	1.3155
	.50	.6680	.3359	.2833	.8471	.1529	.2531	1.5442	2.0883	1.6313	2.9906	.4953	1.6793
	.75	1.0210	.3613	.5018	.8787	.1213	.4537	1.7190	2.4379	2.0502	3.0518	.5259	2.3117
	1.0	1.4144	.4144	.7394	.5712	.4288	.4642	2.6313	4.2625	3.1817	1.6706	.1647	.2084
100	.10	.1270	.2696	.0270	1.0295	.0295	.0295	.4691	.0618	.0309	2.0545	.0273	.0545
	.50	.6661	.3322	.2601	.8906	.1094	.2296	1.2593	1.5187	1.3043	2.5027	.2513	1.1034
	.75	.8991	.1988	.2670	.8579	.1421	.2856	1.4025	1.8050	1.6011	2.1824	.0912	.4047
	1.0	1.1423	.1423	.1294	1.1380	.1380	.6761	.8606	.7213	1.3464	2.3285	.1643	1.1818
500	.10	.0911	.0886	.0089	1.0031	.0031	.0031	.4754	.0492	.0246	2.0253	.0127	.0253
	.50	.5134	.0268	.0556	.9859	.0141	.0747	.5720	.1440	.2481	2.0554	.0277	.2670
	.75	.7837	.0450	.1182	.9642	.0358	.1316	.6270	.2540	.3251	2.0866	.0433	.2380
	1.0	1.0569	.0568	.0334	.9618	.0382	.0390	.6497	.2993	.2059	2.0535	.0268	.0719

Table 5 Interpretation of the Bayes factor (B_{12})

B_{12}	$2 \log B_{12}$	Evidence in favor of H_1
< 1	< 0	Negative (H_2 is accepted)
$[1, 3]$	$[0, 2]$	Weak
$[3, 20]$	$[2, 6]$	Positive
$[20, 150]$	$[6, 10]$	Strong
> 150	> 10	Very strong

values of $2 \log B_{12}^{(1)} = 2.206$, $2 \log B_{12}^{(2)} = 6.471$, $2 \log B_{12}^{(3)} = 1.584$ and $2 \log B_{12}^{(4)} = 2.432$ according to contrast hypothesis testing (i), (ii), (iii) and (iv) given above, indicating “positive,” “strong,” “weak” and “positive” evidence in accepted STN-BS distribution hypothesis; see Table 5.

Table 6 Descriptive statistics for ozone data

Median	Mean	SD	CV	CS	CK	Range	Min.	Max.	n
31.5	42.13	32.99	78.30%	1.21	4.11	167	1	168	116

Table 7 The ML estimates and information criteria based on the BS, SN-BS, SNT-BS, T-BS and STN-BS distributions for the ozone data

Model	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	\hat{v}	$\ell(\hat{\theta})$	AIC	BIC
BS	.982	28.023			– 549.097	1102.194	1107.701
SN-BS	1.270	14.835	1.067		– 545.606	1097.211	1105.472
SNT-BS	1.521	11.318	4.372	1.543	– 541.1762	1090.352	1101.367
T-BS	.807	30.905		7	– 543.394	1092.788	1101.049
STN-BS	1.661	106.899	– 5.956	7.472	– 540.598	1089.196	1100.210

Similar to ozone data, the estimated density functions of the fatigue data including the respected histograms, plotted in Fig. 8, and the PP plots and empirical and theoretical cdf plots given in Figs. 9, 10, 11, 12 and 13 confirm again the appropriateness of the STN-BS distribution.

5 Concluding remarks

In this paper, a flexible class of distributions, called the skew- t -normal Birnbaum–Saunders distributions, based on the Birnbaum–Saunders models is introduced and its several properties are obtained. This skewed distribution extends the skew-normal Birnbaum–Saunders distribution, allowing a better prediction of the extreme percentiles. The parameters estimation via ECM algorithm is proposed, and their performances are evaluated by the Monte Carlo method. The simulation study shows the good performance of these estimators. The utility of this class is illustrated by

Fig. 2 Histograms with density estimates for ozone data

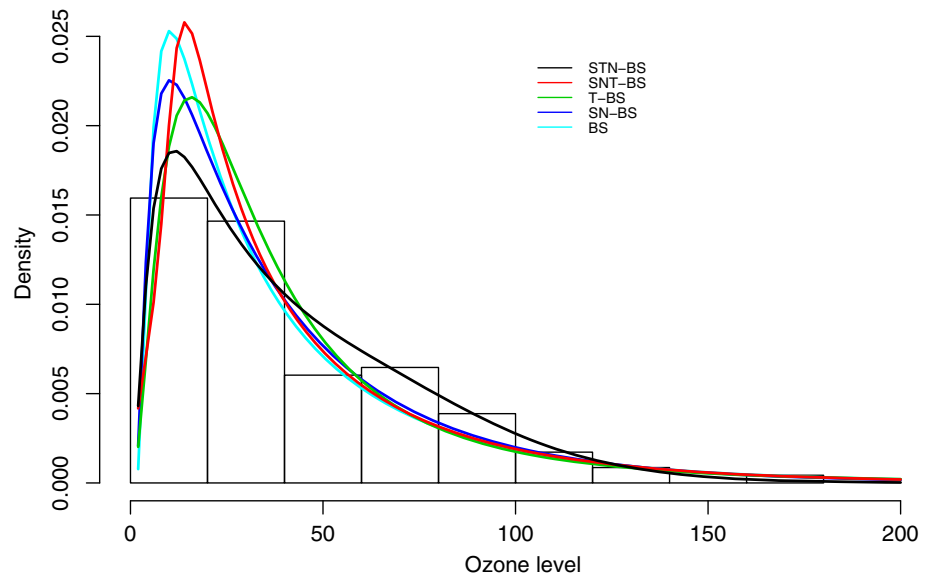


Fig. 3 PP plot (left) and empirical versus theoretical cdf (right) for the BS model of ozone data

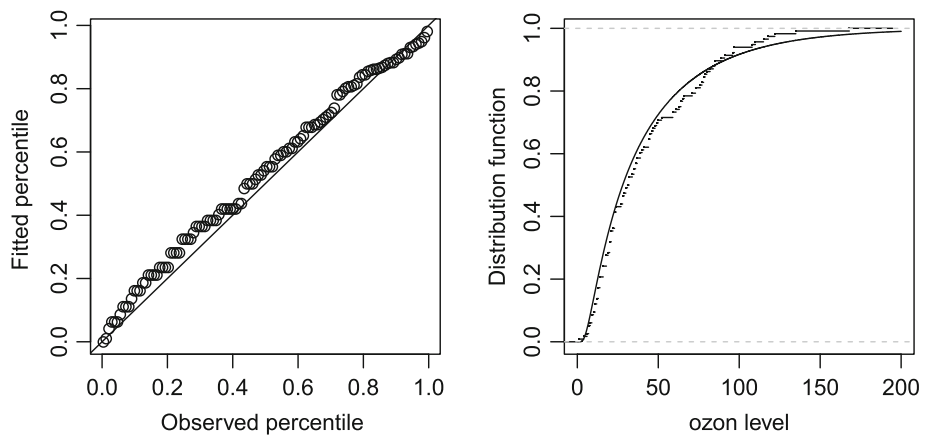
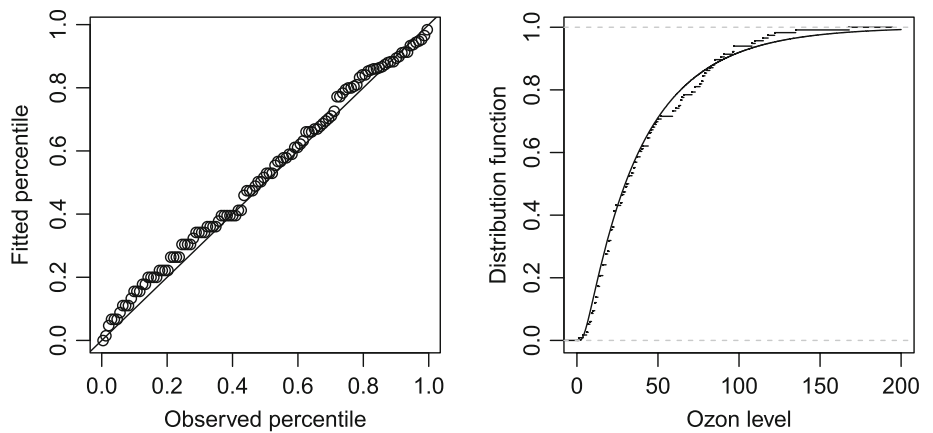


Fig. 4 PP plot (left) and empirical versus theoretical cdf (right) for the SN-BS model of ozone data



means of two real data sets, allowing a better prediction of the extreme percentiles rather than another extensions of the Birnbaum Saunders distribution.

Appendix

This appendix presents the proof of Propositions 2.2 and 2.3.

Fig. 5 PP plot (left) and empirical versus theoretical cdf (right) for the SNT-BS model of ozone data

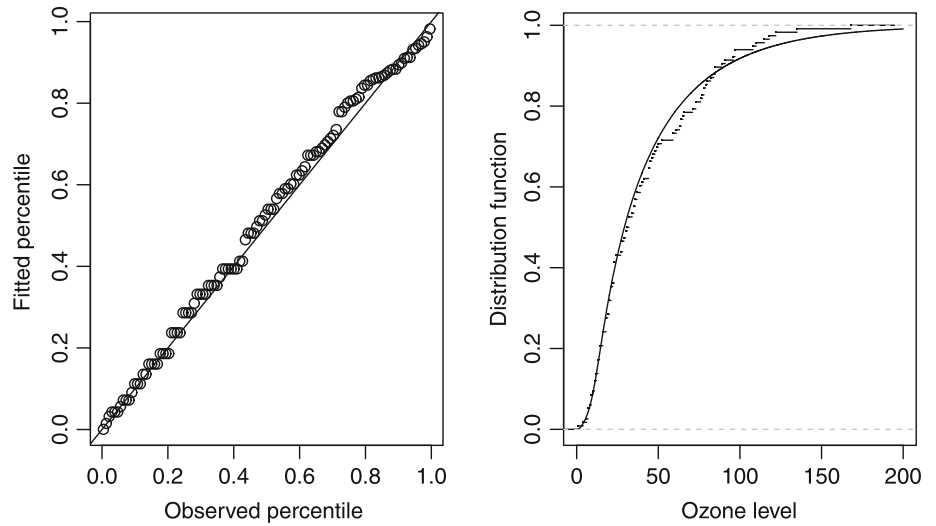


Fig. 6 PP plot (left) and empirical versus theoretical cdf (right) for the T-BS model of ozone data

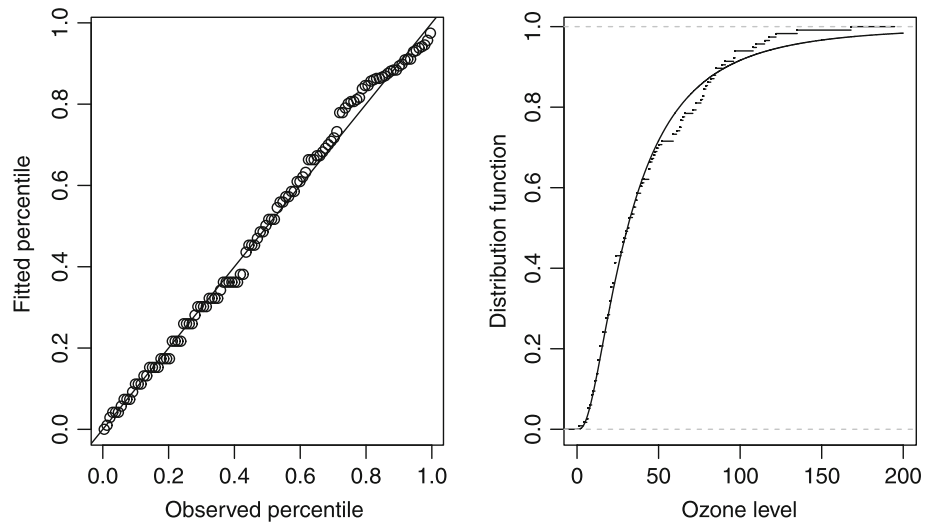


Fig. 7 PP plot (left) and empirical versus theoretical cdf (right) for the STN-BS model of ozone data

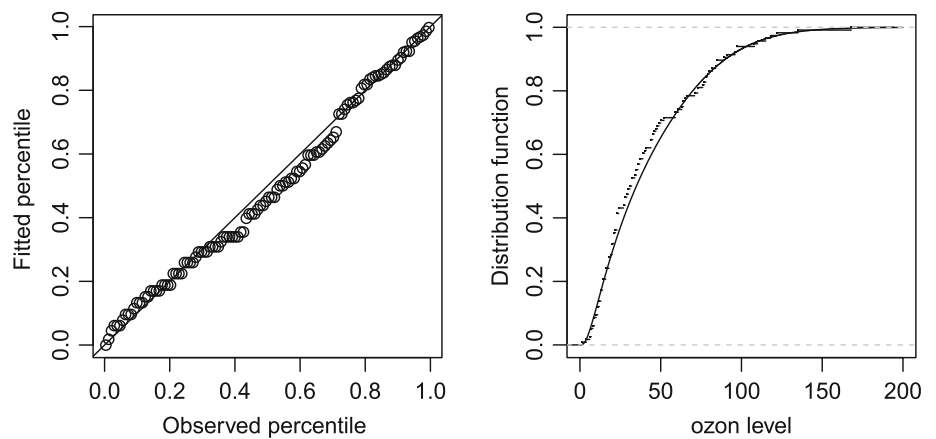


Table 8 Descriptive statistics for fatigue data

Median	Mean	SD	CV	CS	CK	Range	Min.	Max.	<i>n</i>
1416	1401	391.32	.28	.14	2.71	2070	370	2440	101

Table 9 The ML estimates and information criteria based on the BS, SN-BS, SNT-BS, T-BS, STN-BS distribution for the fatigue data

Model	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\nu}$	$\ell(\hat{\theta})$	AIC	BIC
BS	.3110	1335.491			– 751.552	1507.105	1512.335
SN-BS	.320	1239.313	.310		– 751.377	1510.753	1516.598
SNT-BS	.494	1937.772	– 3.515	50	– 746.626	1501.252	1512.266
T-BS	.279	1356.07		13	– 749.263	1504.526	1512.371
STN-BS	.408	1848.170	– 2.298	10.149	– 745.834	1499.669	1510.129

Fig. 8 Histograms with density estimates for fatigue data

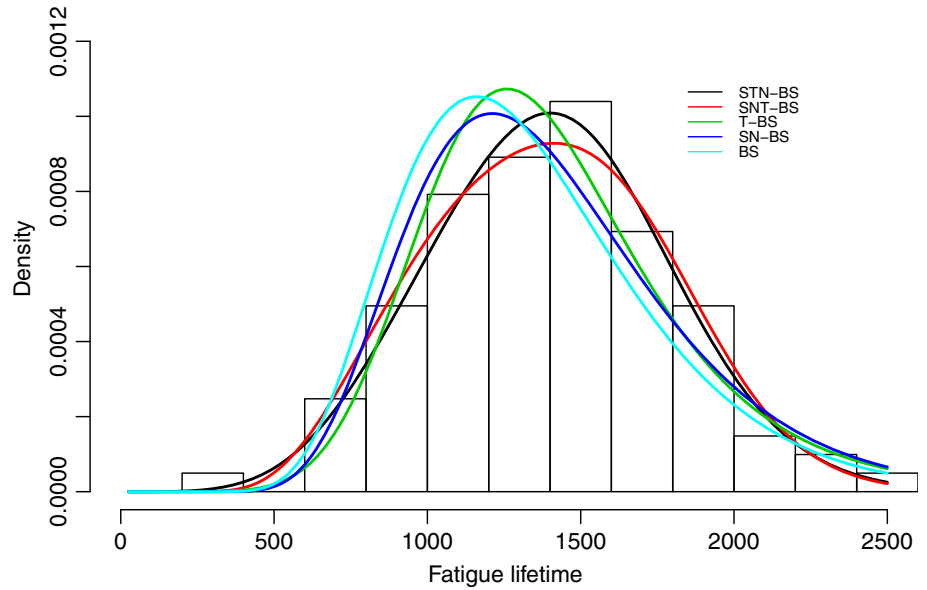
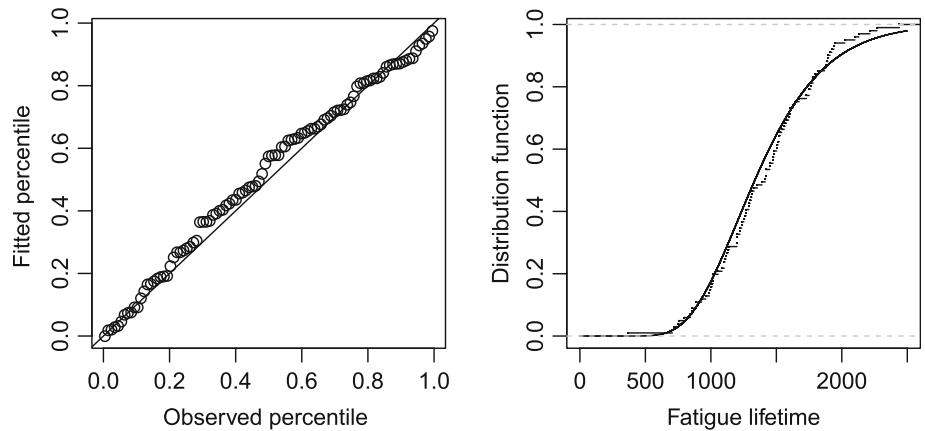


Fig. 9 PP plot (left) and empirical versus theoretical cdf (right) for the BS model of fatigue data



Proof of Proposition 2.2 By using Eq.(5) the cdf of $T | (\gamma, \tau)$ is

$$F(T \leq t | \gamma, \tau) = P(Y \leq a(t; \alpha, \beta) | \gamma, \tau) = P\left(Z_2 \leq \sqrt{\tau + \lambda^2} a(t; \alpha, \beta) - \frac{\lambda\gamma}{\sqrt{\tau + \lambda^2}} | \gamma, \tau\right),$$

then the pdf of $T | \gamma, \tau$ is

$$f(t | \gamma, \tau) = \phi\left(\sqrt{\tau + \lambda^2} a(t; \alpha, \beta) - \frac{\lambda\gamma}{\sqrt{\tau + \lambda^2}}\right) \sqrt{\tau + \lambda^2} A(t; \alpha, \beta).$$

The conditional distribution of $\gamma | \tau$ can be easily obtained from the definitions. \square

Proof of Proposition 2.2 From Proposition 2.2, the joint pdf of T, γ and τ is given by

Fig. 10 PP plot (left) and empirical versus theoretical cdf (right) for the SN-BS model of fatigue data

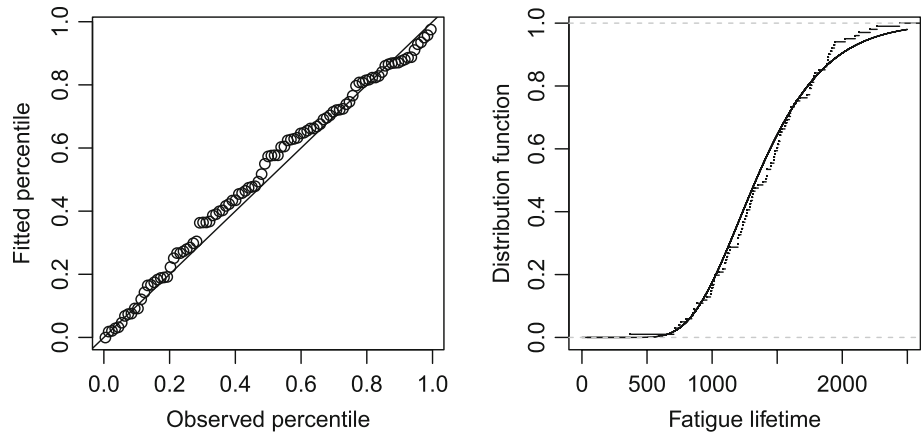


Fig. 11 PP plot (left) and empirical versus theoretical cdf (right) for the SNT-BS model of fatigue data

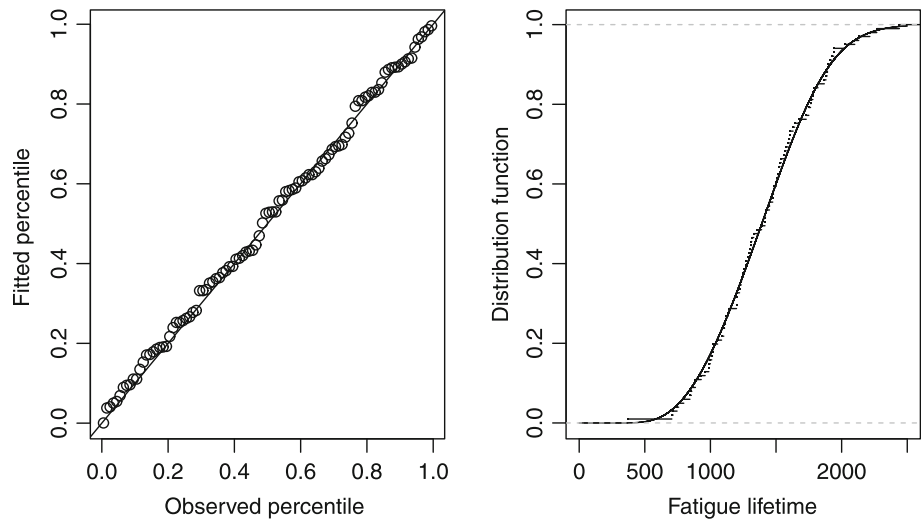


Fig. 12 PP plot (left) and empirical versus theoretical cdf (right) for the T-BS model of fatigue data

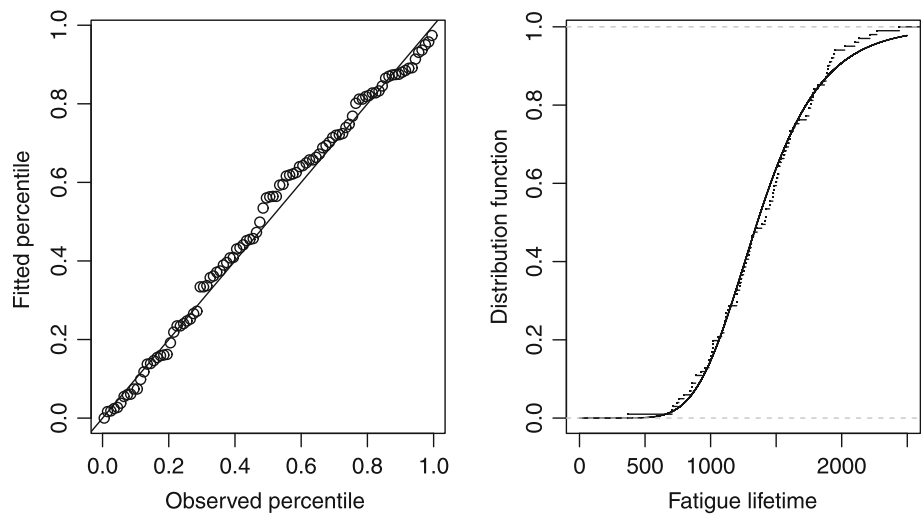
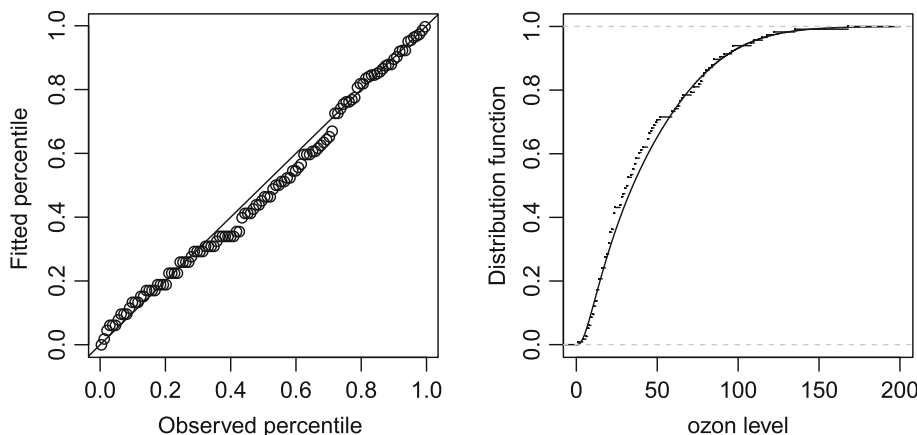


Fig. 13 PP plot (left) and empirical versus theoretical cdf (right) for the STN-BS model of fatigue data



$$f(t, \gamma, \tau) = \frac{1}{\pi} \exp \left\{ -\frac{1}{2} \left(\left(\gamma - \frac{\lambda}{\alpha} \varepsilon(t; \beta) \right)^2 + \frac{\tau}{\alpha^2} \varepsilon^2(t; \beta) + v\tau \right) \right\} \quad (6)$$

$$A(t; \alpha, \beta) \frac{\left(\frac{v}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)} \tau^{\frac{v-1}{2}}$$

where $\varepsilon(t; \beta) = \sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}$.

By integrating on γ in (6), we get

$$f(t, \tau) = \sqrt{\frac{2}{\pi}} \frac{\left(\frac{v}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)} \tau^{\frac{v-1}{2}} A(t; \alpha, \beta) \exp \left\{ \left(-\frac{v + \frac{\varepsilon^2(t; \beta)}{\alpha^2}}{2} \right) \tau \right\} \Phi\left(\lambda \frac{\varepsilon(t; \beta)}{\alpha} \right). \quad (7)$$

Dividing (7) by (3) gives

$$f(\tau | t) = \frac{\left(\frac{v+a^2(t; \alpha, \beta)}{2}\right)^{\frac{v+1}{2}}}{\Gamma\left(\frac{v+1}{2}\right)} \tau^{\frac{v+1}{2}-1} \exp\left(-\frac{v+a^2(t; \alpha, \beta)}{2}\right) \tau,$$

so

$$\tau | (T = t) \sim \Gamma\left(\frac{v+1}{2}, \frac{v+a^2(t; \alpha, \beta)}{2}\right),$$

which concludes parts (a) and (b), and dividing (6) by (7), gives

$$f(\gamma | t, \tau) = \frac{1}{\Phi\left(\lambda \frac{\varepsilon(t; \beta)}{\alpha}\right)} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\gamma - \frac{\lambda}{\alpha} \varepsilon(t; \beta) \right)^2 \right\} I(0, +\infty) = f(\gamma | t),$$

so γ and τ are conditionally independent given $T = t$ and the conditional distribution of γ given t is

$$\gamma | (T = t) \sim TN\left(\lambda a(t; \alpha, \beta), 1, (0, \infty)\right),$$

which concludes the part (c). □

References

Azzalini A (1985) A class of distribution which includes the normal ones. *Scand J Stat* 12:171–178

Azzalini A (1986) Further results on a class of distribution which includes the normal ones. *Statistica* 46:199–208

Barros M, Paula GA, Leiva V (2008) A new class of survival regression models with heavy tailed errors: robustness and diagnostics. *Lifetime Data Anal* 14:316–332

Basford KE, Greenway DR, McLachlan GJ, Peel D (1997) Standard error of fitted means under normal mixture. *Comput Stat* 12:1–17

Birnbaum ZW, Saunders SC (1969) A new family of life distribution. *J Appl Probab* 6:319–327

Cabral CRB, Bolfarine H, Pereira JRG (2008) Bayesian density estimation using skew student-t-normal mixtures. *Comput Stat Data Anal* 52:5075–5090

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J R Stat Soc Ser B* 39:1–38

Desmond A (1985) Stochastic models of failure in random environments. *Can J Stat* 13:171–183

Diaz-Garcia JA, Leiva-Sanchez V (2005) A new family of life distribution based on the elliptically contoured distributions. *J Stat Plan Inference* 128:445–457

Gomez HW, Venegas O, Bolfarine H (2007) Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics* 18:395–407

Gomez HW, Olivares J, Bolfarine H (2009) An extension of the generalized Birnbaum–Saunders distribution. *Stat Probab Lett* 79:331–338

Hashmi F, Amirzadeh V, Jamalizadeh A (2015) An extension of the Birnbaum–Saunders distribution based on skew-normal-t distribution. *Stat Res Train Cent* 12:1–37

Ho HJ, Pyne S, Lin TI (2011) Maximum likelihood inference for mixture of skew Student-t-normal distributions through practical EM-type algorithms. *Stat Comput* 22:287–299

Kass RE, Raftery AE (1995) Bayes factor. *J Am Stat Assoc* 90:773–795

Lachos VH, Dey D, Cancho VG, Louzada N (2017) Scale mixtures log-Birnbaum–Saunders regression models with censored data: a Bayesian approach. *J Stat Comput Simul* 87:2002–2022

Leiva V, Barros M, Paula GA, Galea M (2007) Influence diagnostics in log-Birnbaum–Saunders regression models with censored data. *Comput Stat Data Anal* 51:5694–5707

Leiva V, Riquelme M, Balakrishnan N, Sanhueza A (2008) Lifetime analysis based on the generalized Birnbaum–Saunders. *Comput Stat Data Anal* 52:2079–2097

- Leiva V, Vilca F, Balakrishnan N, Sanhueza A (2010) A skewed sinh-normal distribution and its properties and application to air pollution. *Commun Stat Theory Methods* 39:426–443
- McLachlan GJ, Krishnan T (2008) *The EM algorithm and extensions*, 2nd edn. Wiley, New York
- Meng X-L, Rubin DB (1993) Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* 80:267–278
- Nadarajah S, Kotz S (2003) Skewed distributions generated by the normal kernel. *Stat Probab Lett* 65(3):269–277
- Podaski R (2008) Characterization of diameter data in near-natural forests using the Birnbaum–Saunders distribution. *Can J For Res* 38:518–527
- Raftery AE (1995) Bayesian model selection in social research. *Sociolog Methodol* 25:111–163
- R Development Core Team (2016) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org>
- Sanhueza A, Leiva V, Balakrishnan N (2008) The generalized Birnbaum–Saunders distribution and its methodology and application. *Commun Stat Theory Methods* 37:645–670
- Vilca F, Santana L, Leiva V, Balakrishnan N (2011) Estimation of extreme percentiles in Birnbaum–Saunders distribution. *Comput Stat Data Anal* 55:1665–1678