Chilean Journal of Statistics
Vol. 4, No. 1, April 2013, 87–97

Extreme Value Theory
Research Paper

A study of exponential-type tails applied to Birnbaum-Saunders models

Marta Ferreira

Center of Mathematics of Minho University, Braga, Portugal

(Received: 21 January 2012 · Accepted in final form: 02 January 2013)

Abstract

Birnbaum-Saunders distributions have increasingly been used in environmental sciences applications. A major concern is the adjustment of extreme quantiles. Environmental data have often tails in the Gumbel domain which corresponds to a null tail index and does not allow us to distinguish the different tail weights that might exist between distributions within this domain. Exponential-tail distributions form an important subgroup with the peculiarity of including a parameter that specifies the "penultimate" tail behavior. In particular, we analyze the penultimate tail behavior of Birnbaum-Saunders distributions. We find examples with "heavier" tails than the classical one that can better accommodate environmental data highly concentrated on the right tail. This is illustrated with an application.

Keywords: Exponential-tail models · Extreme value theory · Penultimate approximation.

Mathematics Subject Classification: Primary 60G70 · Secondary 60E05.

1. Introduction

Problems arising in many practical applications have been leading to a major development in the construction of flexible parametric models. Distributions commonly used for modeling environmental data are the so-called life distributions, usually positively skewed, unimodal, and having two parameters; see, for instance, Marshall and Olkin (2007). An example of a life distribution that has been largely studied and applied is the Birnbaum-Saunders (BS) model, which was originated from a problem of material fatigue; see Birnbaum and Saunders (1969). More precisely, the BS distribution relates the total time until the failure to some type of cumulative damage normally distributed. Among many attractive properties, we highlight its relationship with the normal distribution. A random variable (r.v.) having the BS distribution can be represented by another r.v. used as basis. Therefore, by considering different distributions for the basis variable (under diverse arguments) we obtain more classes of models. Even though several generalizations of the BS distribution have been proposed in literature (see Díaz-García and Leiva, 2005; Leiva et al., 2010; Vilca and Leiva, 2006; Vilca et al., 2010, 2011, among others), these models are still inadequate to fit data that are largely concentrated on the left/right tail

Corresponding author. Marta Ferreira. Departamento de Matemática e Aplicações, Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal. Email: msferreira@math.uminho.pt

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)
© Chilean Statistical Society – Sociedad Chilena de Estadística
http://www.soche.cl/chjs

of the distribution. Recently, Ferreira et al. (2012) proposed the EVBS (EVBS*) model, generated from extreme value models, which introduces a tail parameter that allow us to better accommodate the high (low) extreme values. Extreme value models correspond to the limiting class of a suitable linear normalization of maxima (minima). However, there might exist penultimate limiting distributions that better approximate the extreme values; see Fisher and Tippett (1928). This is particularly evident in the Gumbel domain of attraction. Exponential-tail models (ET) include several life distributions and are a wide class in the Gumbel domain. The main purpose of this work is to analyze the penultimate tail behavior of this latter class, so as to achieve such distinction. We focus the particular case of BS models and thus contribute to the identification of models that may be better profiled for environmental data highly concentrated on the right tail.

This paper is organized as follows. In Section 2, we present some basic notions concerning extreme value theory and BS models, with particular emphasis on EVBS distributions. In Section 3, we devote to the analysis of the "penultimate" tail behavior of ET distributions. We study, in particular, the case of BS models. We see that, for instance, the EVBS model generated from Gumbel (Ferreira et al., 2012) has "heavier" tail than the classical BS, and hence it can better accommodate data within the domain of ET distributions that presents observations much concentrated on the right tail. An illustration is provided at the end (Section 4) with an application to real data.

2. Background

In this section we give some preliminary aspects and results of extreme value theory, the classical BS model and some generalizations commonly denoted BS-type models. In particular, we give a brief characterization of the EVBS/EVBS* models.

2.1 Extremal domains of attraction

The distinguishing feature of an extreme value analysis is the development of models and techniques to describe a process at unusual or even unobserved levels. The central result in classical extreme value theory (EVT) states that, for an i.i.d. sequence, $\{X_n\}_{n\geq 1}$, having common cumulative distribution function (c.d.f.) F, if there are real constants $a_n > 0$ and b_n such that,

$$\mathrm{P}(\max(X_1, \ldots, X_n) \leq a_n x + b_n) = F^n(a_n x + b_n) \underset{n\to\infty}{\longrightarrow} G_\gamma(x)\,, \tag{1}$$

for some non-degenerate function G_γ, then it must be the generalized extreme value function (GEV) given by

$$G_\gamma(x) = \exp\left(-\left(1+\gamma x\right)^{-1/\gamma}\right),\ 1+\gamma x > 0,\ \gamma \in \mathrm{R}, \tag{2}$$

($G_0(x) = \exp(-e^{-x})$) and we say that F belongs to the domain of attraction of G_γ, in short, $F \in \mathcal{D}(G_\gamma)$. The parameter γ, known as the tail index, is a shape parameter that determines the tail behavior of F, being so a crucial issue in EVT. More precisely, if $\gamma > 0$, we are in the domain of attraction Fréchet corresponding to a heavy tail, $\gamma < 0$ indicates the Weibull domain of attraction of light tails and $\gamma = 0$ means a Gumbel domain of attraction and an exponential tail. We can rewrite a result similar to the one given in Equation (1) for minima, with a limiting c.d.f. GEV function for minima (GEV*) expressed as

$$G^*_\gamma(x) = 1 - G_\gamma(-x) \tag{3}$$

and say that F belongs to the min-domain of attraction of G^*_γ, in short $F \in \mathcal{D}_{\rm m}(G^*_\gamma)$. This is due to the relation $\min\{X_1, \ldots, X_n\} = -\max\{-X_1, \ldots, -X_n\}$, and hence, in practice, we need to treat only one of the cases, usually the maxima.

Despite the importance of the limiting result in Equation (1), however, as Fisher and Tippett (1928) remarked, if one approximates the distribution of the successive maxima of normal samples not by the limit distribution, which is Gumbel, but by a sequence of other extreme value distributions converging to the limit distribution, the approximation is asymptotically improved. They called penultimate distributions to this sequence of approximating extreme value distributions. This issue have been latter developed by Anderson (1976), Cohen (1982), Gomes (1984, 1993), Canto e Castro (1992), Gomes and de Haan (1999) and Kaufmann (2000), among others. The penultimate tail analysis is particularly important in the Gumbel domain. Observe that the Gumbel domain only comprises distributions having null tail index, i.e., $\gamma = 0$, and a way to distinguish between them in high quantiles is to look at the penultimate tail behavior. ET models are a wide class in the Gumbel domain that allow us such distinction, as will be seen in Section 3.

2.2 Birnbaum-Saunders distributions

An r.v. T with classical BS distribution, denoted by $T \sim \mathrm{BS}(\alpha, \beta)$, is characterized by its shape and scale parameters $\alpha > 0$ and $\beta > 0$, respectively. BS and standard normal r.v.'s, denoted respectively by T and Z, are related by

$$T = \beta\big(\alpha\, Z/2 + \sqrt{(\alpha\, Z/2)^2 + 1}\big)^2 \quad \text{and} \quad Z = \big[\sqrt{T/\beta} - \sqrt{\beta/T}\big]/\alpha. \tag{4}$$

The probability density function (p.d.f.) and c.d.f. of T are, respectively, given by

$$f_T(t) = \phi\big(a(t)\big)\, a'(t) \quad \text{and} \quad F_T(t) = \Phi\big((a(t)\big), \quad t > 0, \tag{5}$$

where ϕ and Φ are the standard normal p.d.f. and c.d.f., respectively,

$$a(t) \equiv a_t = \left[\sqrt{t/\beta} - \sqrt{\beta/t}\right]/\alpha \quad \text{and} \quad a'(t) \equiv A_t = t^{-3/2}\,[t+\beta]\,/\left[2\alpha\sqrt{\beta}\right],$$

where $a'(t) = d\,a(t)/dt$ is the derivative of $a(t)$ with respect to t. The quantile function (q.f.) of T is expressed as

$$t(q) \equiv t_q = F_T^{-1}(q) = \beta\Big(\alpha\,\xi_q/2 + \sqrt{(\alpha\;\xi_q/2)^2 + 1}\Big)^2, \quad 0 < q < 1, \tag{6}$$

where $F_T^{-1}(t) := \inf\{x : F(x) \geq t\}$ is the generalized inverse function of the c.d.f. of T and ξ_q is the qth quantile of the r.v. $Z \sim \mathrm{N}(0, 1)$.

If we switch the standard normal assumption of Z by any other distribution with c.d.f. and p.d.f. F_Z and f_Z, respectively, we obtain the general class of BS-type distributions earlier mentioned, for an associated r.v. T and whose c.d.f. and p.d.f. are, respectively, given by

$$F_T(t) = F_Z\big(a(t)\big) \text{ and } f_T(t) = f_Z\big(a(t)\big)\, a'(t), \quad t > 0. \tag{7}$$

For instance, if Z follows a standard symmetric distribution in the real number set, we then find the GBS distribution, i.e., $T \sim \mathrm{GBS}(\alpha, \beta; g)$, where g is the kernel of the p.d.f. of Z given by $f_Z(z) = c\, g(z^2)$, with $z \in \mathrm{R}$ and c being a normalization constant; see Díaz-García and Leiva (2005). Classical and generalized versions of the BS distribution were implemented in the packages `bs` and `gbs` of the `R` software and are available in `http://CRAN.R-project.org`; see Leiva et al. (2006) and Barros et al. (2009). Among other aspects, these packages include functions for computing probabilities, generating random numbers, estimating parameters and goodness-of-fit analysis.

2.3 EVBS and EVBS* models: properties, extremal domains of attraction and inference

EVBS and EVBS* are BS-type models generated from the extreme value distributions, respectively, GEV and GEV*. More precisely, if Z follows a GEV distribution given in Equation (2), we have the EVBS distribution, $T \sim \text{EVBS}(\alpha,\beta,\gamma)$, where γ is the tail index indicating the domain of attraction of maxima, i.e., the type of the right-tail. Analogously, if Z follows a GEV* distribution given in Equation (3), then $T \sim \text{EVBS}^*(\alpha,\beta,\gamma)$ and γ is the tail index of the left-tail; see Ferreira et al. (2012). Both EVBS and EVBS* models have been analyzed in detail in Ferreira et al. (2012) and Gomes et al. (2012). Some interesting properties concerning scale and inverse operations were proved. Specifically, if $T \sim \text{EVBS}(\alpha,\beta,\gamma)$, then $cT \sim \text{EVBS}(\alpha,c\beta,\gamma)$ and $1/T \sim \text{EVBS}^*(\alpha,1/\beta,\gamma)$, with similar results for EVBS*. It was also shown that, if $Z \in \mathcal{D}(G_\gamma)$, then T belongs to the domain of attraction Fréchet(2γ) or Weibull(γ), whenever $\gamma > 0$ or $\gamma < 0$, respectively. On the other hand, if $Z \in \mathcal{D}(G^*_\gamma)$, then T belongs to the min-domain of attraction Fréchet(-2γ) or Weibull(γ), whenever $\gamma > 0$ or $\gamma < 0$, respectively. In what concerns the Gumbel domain of attraction, it was shown that, in several cases, T belongs to this domain whenever Z belongs too, either for maximum or minimum domains. Moments are highly influenced by the tail heaviness. For example, a GEV(γ) distribution with $\gamma > 1$ do not even have the first moment. This issue was analyzed in Gomes et al. (2012), concerning EVBS and EVBS* models. More precisely, it was proved that the rth moment of T exists if $\mathrm{E}\left[Z^{k+l}\big((\alpha Z)^2+4\big)^{[k-l]/2}\right] < \infty$, with $k = 0,\ldots,r$ and $l = 0,\ldots,k$, and we have

$$\mathrm{E}\left[T^r\right] = \delta^r \sum_{k=0}^{r} \binom{r}{k} \sum_{l=0}^{k} \binom{k}{l} 2^k \mathrm{E}\left[(\alpha Z/2)^{k+l}\left((\alpha Z/2)^2+1\right)^{[k-l]/2}\right].$$

Inference aspects for EVBS distributions were addressed in Ferreira et al. (2012). The estimation was based on maximum-likelihood and, as the system of likelihood equations does not allow to derive explicit estimation formulas for the parameters, a numerical procedure was considered. An `R` package named `evbs` to analyze data from EVBS models was developed, and its "in progress" version is already available through the authors. This package contains, among other useful functions for EVBS models, the maximum likelihood (ML) estimation methodology implemented in Ferreira et al. (2012).

3. ET Distributions and Penultimate Tail Behavior of BS Models

We say that F is an ET model if

$$1 - F(x) \sim \exp(-H(x)),\ x \geq x_0 \geq 0,\ \theta > 0, \tag{8}$$

where $f(x) \sim g(x)$ means $\lim_{x\to\infty} f(x)/g(x) = 1$, and

$$H(x) = x^{1/\theta} l(x) \text{ or } H^{-1}(x) = x^\theta l^*(x), \tag{9}$$

with H^{-1} denoting the generalized inverse of H and functions l and l^* are slowly varying at infinity (i.e., $l(tx)/l(t) \to 1$ as $t \to \infty$ for all $x > 0$ and the same holds for l^*). Functions H and H^{-1} are called regularly varying with indexes, $1/\theta$ and θ, respectively. The class in Equation (8) includes the Weibull-type models (see Gardes and Girard, 2008) and forms an important subgroup within the Gumbel class ($\gamma = 0$), where the tail behavior can then be specified using coefficient θ. More precisely, the parameter θ, here called the ET-coefficient, governs the tail behavior of F, with larger values indicating heavier tails. Examples of ET models include normal ($\theta = 1/2$), Laplace, Gumbel, logistic ($\theta = 1$) and Weibull with the shape parameter γ ($\theta = 1/\gamma$), among others.

In the following, based on a result given in Gomes (1984), we prove that ET models present Fréchet or Weibull penultimate tail behavior whenever $\theta > 1$ or $\theta < 1$, respectively. We remark that the case $\theta = 1$ cannot be treated in a unified way, since we can find penultimate tail indexes with different rates of convergence; see Gomes (1993). Therefore, we assume that $\theta \neq 1$ all over this section. Furthermore, we also assume that, as $x \to \infty$,

$$\tfrac{xl'(x)}{l(x)} \to 0, \ \tfrac{x^2 l''(x)}{l(x)} \to 0, \ \tfrac{x^3 l'''(x)}{l(x)} \to 0. \tag{10}$$

Note that, if $l(x)$ is monotone for $x \geq x_0 > 0$, then $xl'(x)/l(x) \to 0$, as $x \to \infty$. The other conditions in Equation (10) are also satisfied by the most common models.

Consider

$$k(x) := (-\log(-\log(F(x))))'.$$

Observe that

$$-\log(F(x)) = [1 - F(x)][1 + O(1 - F(x))],$$

and considering Equations (8) and (10), we have

$$k(x) = H'(x)[1 + o(1)]. \tag{11}$$

Analogously, we derive

$$k'(x) = H''(x)[1 + o(1)] \text{ and } k''(x) = H'''(x)[1 + o(1)]. \tag{12}$$

Consider $H(x)$ given in Equation (9). We have

$$H'(x) = x^{\theta^{-1}-1} l(x)\Big(\theta^{-1} + \tfrac{xl'(x)}{l(x)}\Big).$$

$$H''(x) = x^{\theta^{-1}-2} l(x)\Big(\theta^{-1}[\theta^{-1} - 1] + 2\theta^{-1}\tfrac{xl'(x)}{l(x)} + \tfrac{x^2 l''(x)}{l(x)}\Big).$$

$$H'''(x) = x^{\theta^{-1}-3} l(x)\Big(\theta^{-1}[\theta^{-1} - 1][\theta^{-1} - 2] + 3\theta^{-1}[\theta^{-1} - 1]\tfrac{xl'(x)}{l(x)} + 3\theta^{-1}\tfrac{x^2 l''(x)}{l(x)} + \tfrac{x^3 l'''(x)}{l(x)}\Big).$$

By Equations (10), (11) and (12), we have that

$$\begin{aligned} xk(x) &= \theta^{-1} x^{\theta^{-1}} l(x)[1 + o(1)]. \\ x^2 k'(x) &= xk(x)[\theta^{-1} - 1]. \\ x^3 k''(x) &= xk(x)[\theta^{-1} - 2][\theta^{-1} - 1]. \end{aligned} \tag{13}$$

Now consider $x^F := \sup\{x : F(x) < 1\}$ and

$$\varphi(t) = (1/k)'(t) = -k'(t)/(k(t))^2.$$

Based on condition,

$$\lim_{t\to x^F} \frac{\varphi'(t)}{k(t)(\varphi(t))^2} = c < \infty, \tag{14}$$

Gomes (1984) derives bounds for

$$(F^n(a_n x + b_n) - G_{\gamma_n}(x))/\gamma_n^2, \tag{15}$$

with $a_n = 1/k(b_n)$ and

$$b_n = H^{-1}(\log(n)),$$

where

$$\gamma(t) = \varphi(H^{-1}(t)) = -\frac{k'(H^{-1}(t))}{k^2(H^{-1}(t))},$$

and the penultimate tail index is given by

$$\gamma_n = \gamma(\log(n)) = -k'(b_n)/k^2(b_n). \tag{16}$$

Proposition 3.1 ET models satisfying Equations (8)-(10) such that $\theta \neq 1$, present Fréchet and Weibull penultimate tail behavior, if $\theta > 1$ and if $\theta < 1$, respectively.

Proof First, observe that $x^F = +\infty$,

$$\varphi'(t) = -\frac{k''(t)k(t) - 2(k'(t))^2}{(k(t))^3}$$

and

$$\frac{\varphi'(t)}{k(t)(\varphi(t))^2} = 2 - \frac{k''(t)k(t)}{(k'(t))^2}.$$

By Equation (13), we obtain

$$\begin{aligned}\lim_{t\to x^F} \frac{k''(t)k(t)}{(k'(t))^2} &= \lim_{t\to x^F} \frac{t^3 k''(t) t k(t)}{(t^2 k'(t))^2} \\ &= \lim_{t\to x^F} \frac{(\theta^{-2} - 3\theta^{-1} + 2 + o(1))(1 + o(1))}{(\theta^{-1} - 1 + o(1))^2} = \frac{(\theta^{-1} - 2)(\theta^{-1} - 1)}{(\theta^{-1} - 1)^2}\end{aligned}$$

and hence

$$\lim_{t\to x^F} \frac{\varphi'(t)}{k(t)(\varphi(t))^2} = \frac{1}{1-\theta}.$$

Thus, condition given in Equation (14) holds ($\theta \neq 1$). Also, by Equation (13), the penultimate tail index, γ_n, in Equation (16) satisfies

$$\gamma_n = -\frac{b_n^2 k'(b_n)}{(b_n k(b_n))^2} \sim \frac{\theta - 1}{\log(n)}.$$

Therefore, we obtain $\gamma_n > 0$ and $\gamma_n < 0$ if, respectively, $\theta > 1$ and $\theta < 1$.

Moreover, the rate of convergence, γ_n^2, in Equation (15) is of order $(\log(n))^{-2}$.

A BS model of ET-type, i.e., with cdf satisfying Equations (8) and (9), is generated also from an ET-type r.v. Z in Equation (4). Next results allow us to relate the penultimate tail behavior of Z and T.

Proposition 3.2 If the r.v. Z in Equation (4) follows an ET model with ET-coefficient θ, then the r.v. T in Equation (4) has ET distribution with ET-coefficient 2θ.

Proof Observe that

$$1 - F_T(t) = 1 - F_Z(a_t) = \exp(-H(a_t)),$$

where $H(a_t) \sim (\sqrt{t/\beta}/\alpha)^{1/\theta} \ell(\sqrt{t/\beta}/\alpha) = t^{1/(2\theta)} \ell^*(t)$, and $\ell^*(t) = (\alpha\sqrt{\beta})^{-1/\theta} \ell(\sqrt{t/\beta}/\alpha)$ is a slowly varying function.

Corollary 3.3 If the r.v. Z in Equation (4) follows an ET model with ET-coefficient θ, then the r.v. T in Equation (4) has ET distribution with Fréchet and Weibull penultimate tail behavior if $\theta > 1/2$ and if $\theta < 1/2$, respectively.

Table 1 presents examples of BS models with the respective ET-coefficients. Observe that the classical BS presents the lightest tail, whilst the EVBS model with $\gamma = 0$, the EVBS* model with $\gamma = -\gamma^* > -2$, as well as the GBS models with logistic and Laplace kernels, belong penultimately to the Fréchet domain.

Table 1. BS models with the respective distribution of the r.v. Z in Equation (4) and the corresponding ET-coefficient θ.

Z	T
N(0,1); $\theta = 1/2$	BS(α, β); $\theta = 1$
Logistic/Laplace kernel g ($f_Z(x) = g(x^2)$); $\theta = 1$	GBS$(\alpha, \beta; g)$; $\theta = 2$
Gumbel; $\theta = 1$	EVBS$(\alpha, \beta, \gamma = 0)$; $\theta = 2$
Weibull (γ^*); $\theta = 1/\gamma^*$	EVBS$^*(\alpha, \beta, \gamma^*)$; $\theta = 2/\gamma^*$

4. An Illustration with Environmental Data

The BS model is appropriated for describing phenomena involving accumulation of some type, as is the case of environmental contamination. We present a real data set from environmental contamination in Chile.

4.1 The data set and an exploratory data analysis

The data correspond to hourly sulfur dioxide (SO2) concentrations (in ppb = ppm × 1,000) observed at one monitoring station located at Providencia zone of Santiago during March in 2002. Table 2 presents a descriptive summary of these data and Figure 1 (left) the respective histogram. This table and histogram indicate a positively skewed distribution. Figure 1 (right) displays the original boxplot, which is constructed for symmetric data, and the adjusted boxplot for asymmetric distributions (for details about the adjusted boxplot, see Hubert and Vandervieren, 2008, and function adjbox of the R package robustbase). The original boxplot shows several atypical observations lying on the right-tail of the distribution of the data. However, when we produce the adjusted boxplot for asymmetric distributions, there are still atypical observations on the right-tail.

Table 2. Descriptive statistics for air data (in ppm × 1,000).

Median	Mean	SD	CV	CS	CK	Range	Min.	Max.	n
2.000	2.765	2.284	82.609%	2.089	6.220	18	1	19	744

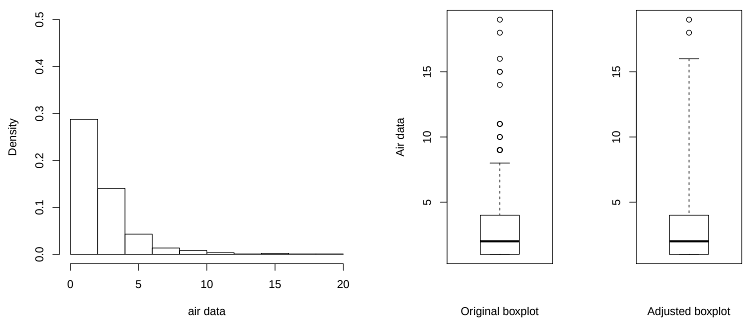

Figure 1. Histogram (left) and indicated boxplot (right) for air data.

4.2 Testing the Gumbel domain of attraction

In order to evaluate if the data belongs to the Gumbel domain of attraction, we carried out a test for the extreme value condition. More precisely, we tested H_0: $F \in \mathcal{D}(G_{\gamma \geq 0})$; details about this test can be seen in Dietrich et al. (2002). Figure 2 (left) presents the sample path of the test statistic as a function of the k largest order statistics. The horizontal line is the critical value above which we reject H_0. Thus, we do not reject the null hypothesis for $1 \lesssim k \lesssim 75$, which is a reasonable value in EV theory to keep it. The hypothesis of a heavy-tail of the Fréchet domain of attraction is removed by Figure 2 (right), corresponding to the sample path of the tail index moments estimator (Dekkers et al., 1989) which is stable around the value zero. Therefore, we have positively skewed data presenting atypical observations on the right-tail but still in the Gumbel domain of attraction. In other words, we are looking for a model within this domain but with a somewhat "heavier" tail, like a penultimate Fréchet behavior. Thus, besides the BS model, we are going to fit to our data the EVBS($\alpha, \beta, \gamma = 0$) model, i.e., the EVBS model based on Gumbel domain which, as already stated, should better accommodate the observations concentrated on the right-tail. For comparison, we also consider the GBS model based on logistic and Laplace kernels.

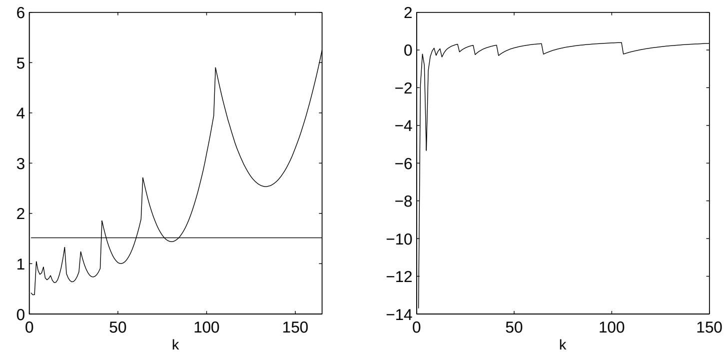

Figure 2. Sample path of the extreme value condition test, where the horizontal line is the critical value above which we reject $F \in \mathcal{D}(G_\gamma)$, with $\gamma \geq 0$ (left) and sample path of the indicated tail index moments estimator (right), for air data.

4.3 Estimation and model checking

Our estimation procedure is based on the maximum likelihood (ML) method. This topic is developed in Ferreira et al. (2012) and Barros et al. (2009), concerning the EVBS model and the GBS model, respectively. In particular, the respective R packages, evbs ("in progress" version available upon request through the authors) and gbs (available in http://CRAN.R-project.org/) were used. Table 3 shows the ML estimates along with the values of the usual information criteria: Akaike (AIC), Schwarz's Bayesian (BIC) and Hannan-Quinn (HQIC) used for model selection. The lower values of AIC, BIC and HQIC in the EVBS model indicate that this is the best of all models, followed by the classical BS model. In addition, Figure 3 confirms this: the solid line corresponding to the EVBS model is always closer to the data than the dotted line corresponding to the classical BS model. Furthermore, a good coherence between the EVBS model and air data is evidenced both by the histogram (left) as by the empirical and theoretical c.d.f.'s (right).

Table 3. ML estimates and information criteria in the indicated models for air data.

Distribution	$\widehat{\alpha}$	$\widehat{\beta}$	$-\ell$	AIC	BIC	HQIC
EVBS$(\alpha, \beta, \gamma = 0)$	0.585	1.569	1303.555	1.755	1.761	1.757
BS(α, β)	0.718	2.205	1336.933	1.800	1.806	1.802
GBS$(\alpha, \beta; g)$(g-logistic kernel)	0.423	2.123	1355.376	1.824	1.831	1.827
GBS$(\alpha, \beta; g)$(g-laplace kernel)	0.584	2.000	1388.753	1.869	1.875	1.872

4.4 Inference on high levels

In the presence of outlying observations, the classical BS may do not do the best job, mainly in the modeling of the large values. This is a very important issue, for instance, in establishing administrative tolerable limiting values in environmental contamination. Considering that the administrative target to abate air pollution usually belongs to the $[98 - 99.9]$ percentile range (Leiva et al., 2008), it is very important to fit well the distribution in this range of higher concentrations. In computing these quantiles for the adjusted BS model (see Table 3), and using formula given in Equation (6), we obtain $[8.6392 - 14.9330]$. Observe that, if this model was adopted to set the administrative target, the values above 15 presented in the box-plot of Figure 1 could not be detected and thus could lead to a wrong decision. Now, if we consider the EVBS distribution generated from Gumbel, which

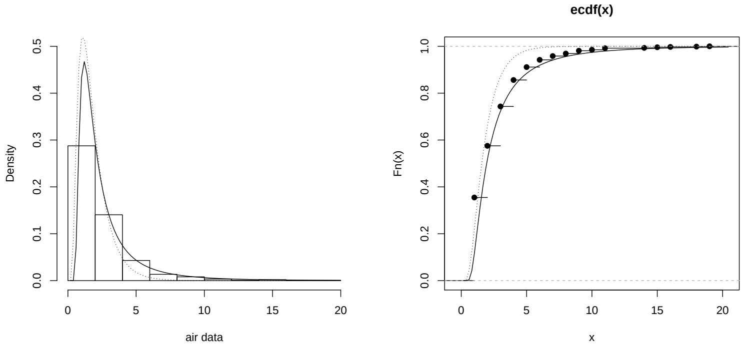

Figure 3. [Left] histogram with estimated EVBS($\widehat{\alpha}, \widehat{\beta}, \gamma = 0$) density (solid line) and estimated classical BS($\widehat{\alpha}, \widehat{\beta}$) density (dotted line); [right] empirical c.d.f. plots with the estimated EVBS($\widehat{\alpha}, \widehat{\beta}, \gamma = 0$) c.d.f. (solid line) and estimated classical BS($\widehat{\alpha}, \widehat{\beta}$) c.d.f. (dotted line), for `air data`.

has shown to be a more suitable model for our data, we obtain $[11.09119 - 28.67016]$ covering all the observed high values. In addition, if we want to infer the probability of exceeding an unusual (but possible) high level, say 30 ppb (see Leiva et al., 2008), using the cdf formulas given in Equations (5) and (7), we will have, respectively, a practically impossible event in the BS model and an estimate of 0.0008 in the EVBS case.

5. Concluding Remarks

Here, we have analyzed the tail behavior of ET distributions and contribute to the study of Birnbaum-Saunders distributions, mainly devoted to environmental phenomena. As already mentioned, the major difficulty within this context concerns the modeling of the tails. Therefore, we think that the extreme value theory has a role to play in Birnbaum-Saunders world, of which ET models are just a part. In the future, we intend to analyze other type of distributions linking extreme value and Birnbaum-Saunders theories, specially concerning the Gumbel domain, the most typical case of environmental data.

Acknowledgements

The author was financed by FEDER Funds through “Programa Operacional Factores de Competitividade - COMPETE”, by Portuguese Funds through FCT - “Fundação para a Ciência e a Tecnologia”, within the Project Est-C/MAT/UI0013/2011 and by the research projects PTDC/MAT/101736/2008 and PTDC/MAT/112770/2009. The author wishes to thank the referees for their valuable contributions to the final form of this paper and Professor Victor Leiva (Valparaíso University, Chile) for providing the data.

References

Anderson, C.W., 1976. Extreme value theory and its approximations. Proceedings Syrup. Reliability Technology, Bradford.

Barros, M., Paula, G.A., Leiva, V., 2009. An R implementation for generalized Birnbaum-Saunders distributions. Computational Statistics and Data Analysis, 53, 1511–1528.

Birnbaum, Z.W., Saunders, S.C., 1969. A new family of life distributions. Journal of Applied Probability, 6, 319–327.

Canto e Castro, L., 1992. Sobre a teoria assintótica de extremos. Ph.D. Thesis. Faculdade de Ciências da Universidade de Lisboa, Portugal.
Cohen, J.P., 1982. Convergence rates for the ultimate and penultimate approximations in extreme-value theory. Advances in Applied Probability, 14, 833–854.
Dekkers, A.L., Einmahl, J., De Haan, L., 1989. A moment estimator for the index of an extreme-value distribution. The Annals of Statistics, 17, 1833–1855.
Díaz-García, J.A., Leiva, V., 2005. A new family of life distributions based on elliptically contoured distributions. Journal of Statistical Planning and Inference, 128, 445–457.
Dietrich, D., de Haan, L., Hüsler, J., 2002. Testing extreme value conditions. Extremes, 5, 71–85.
Ferreira, M., Gomes, M.I., Leiva, V., 2012. On an extreme value version of the Birnbaum-Saunders distribution. Revstat, 10, 181–210.
Fisher, R.A., Tippett, L.H.C., 1928. Limiting forms of the frequency distribution of the largest and smallest member of a sample. Proceedings of the Cambridge Philosophical Society, 24, 180–190.
Gardes, L., Girard, S., 2008. Estimation of the Weibull tail-coefficient with linear combination of upper order statistics. Journal of Statistical Planning and Inference, 138, 1416–1427.
Gomes, M.I., 1984. Penultimate limiting forms in extreme value theory. Annals of the Institute of Statistical Mathematics, 36, 71–85.
Gomes, M.I., 1993. Penultimate behaviour of the extremes. Proceedings of the Conference on Extreme Value Theory and its Applications, Gaithersburg, pp. 40–418.
Gomes, M.I., de Haan, L., 1999. Approximation by penultimate extreme value distributions. Extremes, 2, 71–85.
Gomes, M.I., Ferreira, M., Leiva, V., 2012. The extreme value Birnbaum-Saunders model, its moments and an application in Biometry. Biometrical Letters, 49, 81–94.
Hubert, M., Vandervieren, E., 2008. An adjusted boxplot for skewed distributions. Computational Statistics and Data Analysis, 52, 5186–5201.
Kaufmann, E., 2000. Penultimate approximations in extreme value theory. Extremes, 3, 39–55.
Leiva, V., Barros, M., Paula, G.A., Sanhueza, A., 2008. Generalized Birnbaum-Saunders distributions applied to air pollutant concentration. Environmetrics, 19, 235–249.
Leiva, V., Hernández, H., Riquelme, M., 2006. A new package for the Birnbaum-Saunders distribution. R Journal, 6, 35–40.
Leiva, V., Vilca, F., Balakrishnan, N., Sanhueza, A., 2010. A skewed sinh-normal distribution and its properties and application to air pollution. Communications in Statistics - Theory and Methods, 39, 426–443.
Marshall, A.W., Olkin, I., 2007. Life Distributions. Springer, New York.
Vilca, F., Leiva, V., 2006. A new fatigue life model based on the family of skew-elliptical distributions. Communications in Statistics - Theory and Methods, 35, 229–244.
Vilca, F., Sanhueza, A., Leiva, V., Christakos, G., 2010. An extended Birnbaum-Saunders model and its application in the study of environmental quality in Santiago, Chile. Stochastic Environmental Research and Risk Assessment, 24, 771–782.
Vilca, F., Santana, L., Leiva, V., Balakrishnan, N., 2011. Estimation of extreme percentiles in Birnbaum-Saunders distributions. Computational Statistics and Data Analysis, 55, 1665–1678.