Brazilian Journal of Probability and Statistics
2018, Vol. 32, No. 1, 69–85
https://doi.org/10.1214/16-BJPS332

Improved inference for the generalized Pareto distribution

Juliana F. Pires[a], **Audrey H. M. A. Cysneiros**[b] **and Francisco Cribari-Neto**[b]

[a]*Universidade Federal da Paraíba*
[b]*Universidade Federal de Pernambuco*

Abstract. The generalized Pareto distribution is commonly used to model exceedances over a threshold. In this paper, we obtain adjustments to the generalized Pareto profile likelihood function using the likelihood function modifications proposed by Barndorff-Nielsen (*Biometrika* **70** (1983) 343–365), Cox and Reid (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **55** (1993) 467–471), Fraser and Reid (*Utilitas Mathematica* **47** (1995) 33–53), Fraser, Reid and Wu (*Biometrika* **86** (1999) 249–264) and Severini (*Biometrika* **86** (1999) 235–247). We consider inference on the generalized Pareto distribution shape parameter, the scale parameter being a nuisance parameter. Bootstrap-based testing inference is also considered. Monte Carlo simulation results on the finite sample performances of the usual profile maximum likelihood estimator and profile likelihood ratio test and also their modified versions is presented and discussed. The numerical evidence favors the modified profile maximum likelihood estimators and tests we propose. Finally, we consider two real datasets as illustrations.

1 Introduction

Extreme value analysis is widely used in several fields such as reliability, insurance, engineering and environment (Coles, 2001, Castillo et al., 2005). The standard approach in the analysis of extreme values is based on inferences on the generalized extreme value distribution, which is suitable when the data consists of a set of maxima. However, in some practical applications, when data on maxima of sets of observations are used important information contained in the observations that are not maximal is neglected. For example, in designing a dam, engineers are typically interested in the maximal precipitation, but they may also be interested in all amounts of precipitation which exceed the dam storage capability. In such situations, more information can be gathered by using all observations that exceed a given threshold. The differences between these values and a given threshold are called exceedances over the threshold. These exceedances are typically modeled by the generalized Pareto distribution (GPD).

Since the introduction of the GPD by Pickands (1975) several methods have been considered in the literature for making inferences on its parameters. Estimators derived using the method of moments and the method of probability-weighted

Key words and phrases. Bootstrap, generalized Pareto distribution, likelihood ratio test, maximum likelihood estimation, profile likelihood.

Received March 2015; accepted August 2016.

moments are available in Hosking and Wallis (1987). Castillo and Hadi (1997) proposed the percentile estimation method. Recently, Giles, Feng and Godwin (2011) used the results in Cox and Snell (1968) to obtain bias-corrected maximum likelihood estimators of the parameters that index the GPD.

Davison (2003) used the GPD to model Danish fire insurance claim exceedances data. de Carvalho, Turkman and Rua (2013) proposed a GPD Box–Jenkins-like model, the Box–Jenkins–Pareto model, and used it to analyze weekly unemployment insurance claims in the USA. A generalization of the GPD exceedances over thresholds concept to the space of continuous functions was considered by Ferreira and Haan (2014) and used to simulate wind fields connected to disastrous storms on the basis of observed extreme but not disastrous storms.

Practitioners are frequently interested in performing inference on one of the parameters that index the model, the remaining parameters being nuisance parameters. Such inferences regarding the parameter of interest are commonly based on the profile likelihood function. It is important to note that this function is not a genuine likelihood function and that the null distribution of likelihood ratio test statistic can be poorly approximated by the χ^2 distribution in finite samples. Several authors have proposed modifications that can be applied to the profile likelihood function in order to attenuate such problems; see Barndorff-Nielsen (1983), Cox and Reid (1987, 1993), Fraser and Reid (1995), Fraser, Reid and Wu (1999) and Severini (1999).

Our main goal in this paper is to obtain adjustments to the profile likelihood function that deliver more accurate inferences on the GPD shape parameter. We focus on maximum likelihood estimation and likelihood ratio testing. We also consider bootstrap-based testing inference.

The remainder of the paper unfolds as follows. In Section 2, we review some adjustments to profile likelihood functions. In Section 3, we obtain adjustments to the generalized Pareto log-likelihood function and describe bootstrap-based inference. Monte Carlo simulation results are presented and discussed in Section 4. Two illustrations are presented and discussed in Section 5. Finally, some concluding remarks are given in Section 6.

2 Adjustments to the profile likelihood function

Let y be an $n \times 1$ vector of independent observations from a distribution indexed by the parameter vector θ. Also, let $L(\theta)$ denote the usual likelihood function. Suppose that θ is partitioned as $\theta = (\psi^\top, \lambda^\top)^\top$, where ψ is the parameter of interest and λ is a nuisance parameter; both ψ and λ may be vectors or scalars. Inferences on the parameter of interest may be based on a marginal or on a conditional likelihood function. There are, however, a number of situations in which such functions cannot be explicitly derived. In such cases, the profile likelihood function can be used. The profile likelihood function for ψ, defined as $L_p(\psi) = L(\psi, \widehat{\lambda}_\psi)$, is obtained by replacing λ in the likelihood function by its constrained maximum likelihood estimator $\widehat{\lambda}_\psi$, that is, the maximum likelihood estimator of λ for a given value of ψ.

It is possible to make inferences on ψ using $L_p(\psi)$. The maximum likelihood estimator of ψ, $\widehat{\psi}$, is the value of ψ that maximizes $L_p(\psi)$. The likelihood ratio statistic for testing $\mathcal{H}_0 : \psi = \psi_0$ against $\mathcal{H}_1 : \psi \neq \psi_0$ can be expressed as

$$LR = 2\{\ell_p(\widehat{\psi}) - \ell_p(\psi_0)\},$$

where $\ell_p(\psi) = \log L_p(\psi)$ is the profile log-likelihood function. It is noteworthy, however, that $L_p(\psi)$ is not a genuine likelihood function. For instance, the profile score and information biases are only guaranteed to be of order $\mathcal{O}(1)$.

The profile likelihood function may yield a poor approximation to the true likelihood function in some cases and, as a consequence, the associated estimators and tests may display poor finite sample performances. Several modifications to the profile likelihood function have been proposed in the literature.

Barndorff-Nielsen (1983) proposed a modified profile likelihood function that approximates the marginal or conditional likelihood function for ψ, if either exists. His approach uses the p^* formula, which yields an approximation to the conditional density of $\widehat{\psi}$ given an ancillary statistic. His modified profile log-likelihood function is invariant under interest-respecting reparameterizations and can be expressed as

$$\ell_{BN}(\psi) = \ell_p(\psi) - \log\left|\frac{\partial\widehat{\lambda}_\psi}{\partial\widehat{\lambda}}\right| - \frac{1}{2}\log\left|j_{\lambda\lambda}(\psi, \widehat{\lambda}_\psi; \widehat{\psi}, \widehat{\lambda}, a)\right|, \tag{1}$$

where $j_{\lambda\lambda}(\psi, \lambda; a) = -\partial^2\ell/\partial\lambda\,\partial\lambda^\top$ is the observed information for λ and $\partial\widehat{\lambda}_\psi/\partial\widehat{\lambda}$ is the matrix of partial derivatives of $\widehat{\lambda}_\psi$ with respect to $\widehat{\lambda}$. It is possible to show that the score and information biases are of order $\mathcal{O}(n^{-1})$. The corresponding modified maximum likelihood estimator shall be denoted by $\widehat{\psi}_{BN}$.

There is an alternative expression for the log-likelihood function given in (1) that does not involve $|\partial\widehat{\lambda}_\psi/\partial\widehat{\lambda}|$, namely

$$\ell_{BN}(\psi) = \ell_p(\psi) + \frac{1}{2}\log\left|j_{\lambda\lambda}(\psi, \widehat{\lambda}_\psi; \widehat{\psi}, \widehat{\lambda}, a)\right| - \log\left|\ell_{\lambda;\widehat{\lambda}}(\psi, \widehat{\lambda}_\psi; \widehat{\psi}, \widehat{\lambda}, a)\right|, \tag{2}$$

where $\ell_{\lambda;\widehat{\lambda}}(\psi, \widehat{\lambda}_\psi; \widehat{\psi}, \widehat{\lambda}, a) = \partial^2\ell(\psi, \widehat{\lambda}_\psi; \widehat{\psi}, \widehat{\lambda}, a)/\partial\lambda\,\partial\widehat{\lambda}$. Notice that it involves a sample space derivative and requires the specification of an ancillary a such that $(\widehat{\psi}, \widehat{\lambda}, a)$ is a minimal sufficient statistic.

Some approximations to the modified profile log-likelihood function given in (1) have been proposed in the literature. An approximation to $\ell_{\lambda;\widehat{\lambda}}(\psi, \widehat{\lambda}_\psi; \widehat{\psi}, \widehat{\lambda}, a)$ based on an approximately ancillary statistic was obtained by Fraser and Reid (1995) and Fraser, Reid and Wu (1999):

$$\ell_{FR}(\psi) = \ell_p(\psi) + \frac{1}{2}\log\left|j_{\lambda\lambda}(\psi, \widehat{\lambda}_\psi)\right| - \log\left|\ell_{\lambda;y}(\psi, \widehat{\lambda}_\psi)\widehat{V}_\lambda\right|, \tag{3}$$

where $\ell_{\lambda,y}(\psi,\lambda) = \partial l_\lambda/\partial y^\top$, l_λ being the score function for λ,

$$\widehat{V}_\lambda = \left(-\frac{\partial F(y_1;\widehat{\psi},\widehat{\lambda})/\partial\widehat{\lambda}}{f(y_1;\widehat{\psi},\widehat{\lambda})},\ldots,-\frac{\partial F(y_n;\widehat{\psi},\widehat{\lambda})/\partial\widehat{\lambda}}{f(y_n;\widehat{\psi},\widehat{\lambda})}\right)^\top.$$

Here, $f(y_j;\psi,\lambda)$ is the probability density function of y_j and $F(y_j;\psi,\lambda)$ is the cumulative distribution function of y_j, $j = 1,\ldots,n$. We denote the corresponding modified maximum likelihood estimator by $\widehat{\psi}_{FR}$.

An approximation to Barndorff-Nielsen's modified profile likelihood which can be more easily computed was proposed by Severini (1999). It is based on empirical covariances and is especially useful when expected values of products of log-likelihood derivatives are difficult to obtain. Such an approximation is given by

$$\ell_S(\psi) = \ell_p(\psi) + \frac{1}{2}\log\left|j_{\lambda\lambda}(\psi,\widehat{\lambda}_\psi)\right| - \log\left|\breve{I}_\lambda(\psi,\widehat{\lambda}_\psi;\widehat{\psi},\widehat{\lambda})\right|,$$

where

$$\breve{I}_\lambda(\psi,\lambda;\psi_0,\lambda_0) = \sum_{j=1}^{n}\ell_\lambda^{(j)}(\psi,\lambda)\ell_\lambda^{(j)}(\psi_0,\lambda_0)^\top.$$

Here, $\ell_{\lambda(j)}$ is the score function for the jth observation. The corresponding modified maximum likelihood estimator is denoted by $\widehat{\psi}_S$.

A different adjustment to the profile likelihood function was proposed by Cox and Reid (1987). Their modified profile log-likelihood function is given by

$$\ell_{CR}(\psi) = \ell_p(\psi) - \frac{1}{2}\log\left|j_{\lambda\lambda}(\psi,\widehat{\lambda}_\psi)\right|.$$

The maximizer of $\ell_{CR}(\psi)$ shall be denoted as $\widehat{\psi}_{CR}$. The corresponding score bias is $\mathcal{O}(n^{-1})$ but, in general, the information bias remains of order $\mathcal{O}(1)$. There are two drawbacks associated with this adjustment: it requires an orthogonal parameterization and is not invariant under reparameterizations. An approximation to the adjustment proposed by Cox and Reid (1987) that does not require orthogonality was derived in Cox and Reid (1993) for a scalar parameter of interest. It can be written as

$$\ell_{CR_a}(\psi) = \ell_p(\psi) - \frac{1}{2}\log\left|j_{\lambda\lambda}(\psi,\widehat{\lambda}_\psi)\right| - (\psi-\widehat{\psi})m_\lambda(\widehat{\psi},\widehat{\lambda}),$$

where $m_\lambda(\widehat{\psi},\widehat{\lambda})$ denotes the trace of $\partial m/\partial\lambda$. Here, $m = i^{\lambda\lambda}i_{\psi\lambda}$ and $i_{\psi\lambda} = -E(\partial^2\ell(\psi,\lambda)/\partial\psi\,\partial\lambda)$, with $i^{\lambda\lambda}$ denoting the inverse of $i_{\lambda\lambda} = -E(\partial^2\ell(\psi,\lambda)/\partial\lambda^2)$. This expression obtained by Cox and Reid (1993) is not invariant under reparameterization. The corresponding modified profile maximum likelihood estimator of ψ is denoted by $\widehat{\psi}_{CR_a}$.

For a detailed review of modified profile likelihood functions, see Pace and Salvan (1997) and Severini (2000).

3 Improved inference for the generalized Pareto distribution

Let Y be a random variable following the GPD with a scale parameter $\sigma > 0$ and a shape parameter $-\infty < \xi < \infty$. The distribution and density functions are given, respectively, by

$$F(y) = \begin{cases} 1 - \left(1 + \dfrac{\xi y}{\sigma}\right)^{-\frac{1}{\xi}}, & \xi \neq 0, \\ 1 - \exp\left(-\dfrac{y}{\sigma}\right), & \xi = 0, \end{cases}$$

and

$$f(y) = \begin{cases} \dfrac{1}{\sigma}\left(1 + \dfrac{\xi y}{\sigma}\right)^{-\frac{1}{\xi}-1}, & \xi \neq 0, \\ \dfrac{1}{\sigma}\exp\left(-\dfrac{y}{\sigma}\right), & \xi = 0. \end{cases}$$

The range of y is $0 \leq y < \infty$ if $\xi \geq 0$ and $0 \leq y < -\sigma/\xi$ if $\xi < 0$. The uniform, exponential and Pareto distributions are special cases of the GPD.

The rth central moment of Y is

$$E(Y^r) = \frac{r!\sigma^r}{\prod_{i=1}^{r}(1 - i\xi)}, \qquad r = 1, 2, \ldots.$$

It exists if $\xi < 1/r$.

Let $y = (y_1, \ldots, y_n)^\top$ be a random sample of size n from the GPD and let $\theta = (\xi, \sigma)^\top$ be the unknown parameter vector. The log-likelihood function is given by

$$\ell(\theta) = -n\log(\sigma) - \left(1 + \frac{1}{\xi}\right)\sum_{j=1}^{n}\log\left(1 + \frac{\xi y_j}{\sigma}\right). \tag{4}$$

The score functions for ξ and σ are

$$l_\xi = \frac{1}{\xi^2}\sum_{j=1}^{n}\log\left(1 + \frac{\xi y_j}{\sigma}\right) - \left(1 + \frac{1}{\xi}\right)\sum_{j=1}^{n}\left(\frac{y_j}{\sigma + \xi y_j}\right)$$

and

$$l_\sigma = -\frac{n}{\sigma} + \frac{1+\xi}{\sigma}\sum_{j=1}^{n}\left(\frac{y_j}{\sigma + \xi y_j}\right),$$

respectively.

As discussed in Castillo and Hadi (1997), the maximum likelihood estimators of the GPD parameters may not exist in some cases. In particular, the maximum likelihood estimators of ξ and σ do not exist when $\xi < -1$.

Let ξ be the parameter of interest and σ the nuisance parameter. The restricted maximum likelihood estimator of the parameter σ, $\widehat{\sigma}_\xi$, is obtained by maximizing the log-likelihood function in (4) for fixed ξ. The estimator $\widehat{\sigma}_\xi$ does not have a closed-form expression. It may be found by using constrained nonlinear optimization methods; see, for example, Nocedal and Wright (2006).

It follows from the definition of the profile likelihood function that

$$\ell_p(\xi) = -n\log(\widehat{\sigma}_\xi) - \left(1+\frac{1}{\xi}\right)\sum_{j=1}^{n}\log\left(1+\frac{\xi y_j}{\widehat{\sigma}_\xi}\right). \tag{5}$$

The profile maximum likelihood estimator of ξ, the maximizer of $\ell_p(\xi)$, is denoted by $\widehat{\xi}_p$.

We shall now obtain the modified profile log-likelihoods described in Section 2. In what follows, we shall derive two approximations to the modified profile log-likelihood function developed by Barndorff-Nielsen (1983). The first was proposed by Fraser and Reid (1995) and by Fraser, Reid and Wu (1999). For the GPD, it is given by

$$\ell_{FR}(\xi) = \ell_p(\xi) + \frac{1}{2}\log\left|j_{\sigma\sigma}(\xi,\widehat{\sigma}_\xi)\right| - \log\left|\ell_{\sigma;Y}(\xi,\widehat{\sigma}_\xi)\widehat{V}_\sigma\right|, \tag{6}$$

where

$$j_{\sigma\sigma}(\xi,\widehat{\sigma}_\xi) = -\frac{n}{\widehat{\sigma}_\xi^2} + \frac{1+\xi}{\widehat{\sigma}_\xi^2}\sum_{j=1}^{n}\left(\frac{y_j(2\widehat{\sigma}_\xi+\xi y_j)}{(\widehat{\sigma}_\xi+\xi y_j)^2}\right), \tag{7}$$

and, since $\ell_{\sigma;Y}(\xi,\widehat{\sigma}_\xi) = (\frac{1+\xi}{(\widehat{\sigma}_\xi+\xi y_1)^2},\dots,\frac{1+\xi}{(\widehat{\sigma}_\xi+\xi y_n)^2})$ and $\widehat{V}_\sigma = (\frac{y_1}{\widehat{\sigma}_\xi},\dots,\frac{y_n}{\widehat{\sigma}_\xi})^\top$, we have that,

$$\ell_{\sigma;Y}(\xi,\widehat{\sigma}_\xi)\widehat{V}_\sigma = \sum_{j=1}^{n}\left(\frac{y_j(1+\xi)}{\widehat{\sigma}(\widehat{\sigma}_\xi+\xi y_j)^2}\right).$$

The second approximation follows from the results in Severini (1999) and can be expressed as

$$\ell_S(\xi) = \ell_p(\xi) + \frac{1}{2}\log\left|j_{\sigma\sigma}(\xi,\widehat{\sigma}_\xi)\right| - \log\left|I_\sigma(\xi,\widehat{\sigma}_\xi;\widehat{\xi},\widehat{\sigma})\right|, \tag{8}$$

where

$$I_\sigma(\xi,\widehat{\sigma}_\xi;\widehat{\xi},\widehat{\sigma}) = \sum_{j=1}^{n}\left[-\frac{1}{\widehat{\sigma}_\xi} + \frac{y_j(1+\xi)}{\widehat{\sigma}_\xi(\widehat{\sigma}_\xi+\xi y_j)}\right]\left[-\frac{1}{\widehat{\sigma}} + \frac{y_j(1+\widehat{\xi})}{\widehat{\sigma}(\widehat{\sigma}+\widehat{\xi} y_j)}\right].$$

The modified profile log-likelihood function proposed by Cox and Reid (1987) requires parameter orthogonality, but ξ and σ are not orthogonal, so the adjustment proposed by Cox and Reid (1993) is more appropriate. Applying Cox and Reid (1993, page 469, equation (7)) for the GPD, we obtain

$$\ell_{CR_a}(\xi) = \ell_p(\xi) - \frac{1}{2}\log\left|j_{\sigma\sigma}(\xi, \widehat{\sigma}_\xi)\right| + \xi m_\sigma(\widehat{\xi}, \widehat{\sigma}), \tag{9}$$

where $j_{\sigma\sigma}$ is as given in (7) and $m_\sigma(\widehat{\xi}, \widehat{\sigma}) = \partial m/\partial\sigma$ with $m = i^{\sigma\sigma} i_{\xi\sigma}$. Since in our case $i^{\sigma\sigma} = \sigma^2(1+2\xi)/n$ and $i_{\xi\sigma} = n/[\sigma(1+\xi)(1+2\xi)]$, then, $m_\sigma(\widehat{\xi}, \widehat{\sigma}) = 1/(1+\widehat{\xi})$.

The modified profile maximum likelihood estimators of ξ obtained from (6), (8) and (9) are denoted, respectively, by $\widehat{\xi}_{FR}$, $\widehat{\xi}_S$ and $\widehat{\xi}_{CR_a}$. These estimators and the profile maximum likelihood estimator $\widehat{\xi}_p$ cannot be expressed in closed-form. They are computed by numerically maximizing the corresponding modified profile log-likelihood functions.

The likelihood ratio statistics obtained from the profile and modified profile log-likelihood functions given above for the test of $\mathcal{H}_0 : \xi = \xi_0$ against $\mathcal{H}_1 : \xi \neq \xi_0$ are

$$\begin{aligned} LR_p &= 2\{\ell_p(\widehat{\xi}_p) - \ell_p(\xi_0)\}, \\ LR_{FR} &= 2\{\ell_{FR}(\widehat{\xi}_{FR}) - \ell_{FR}(\xi_0)\}, \\ LR_S &= 2\{\ell_S(\widehat{\xi}_S) - \ell_S(\xi_0)\} \end{aligned}$$

and

$$LR_{CR_a} = 2\{\ell_{CR_a}(\widehat{\xi}_{CR_a}) - \ell_{CR_a}(\xi_0)\}.$$

Under some regularity conditions (see Cox and Hinkley, 1974) and the null hypothesis, they are all asymptotically distributed as χ^2_1.

Hypothesis testing can also be performed using bootstrap resampling (Efron, 1979). Here, inference is based on the comparison between the test statistic computed using the original sample and a critical value obtained from an estimate of its null distribution constructed using a set of artificial (computer generated) samples. In particular, B pseudo-samples of size n are generated from the GPD under the null hypothesis and the test statistic LR_p^{*b} is computed for each pseudo-sample, $b = 1, \dots, B$. The critical value (say, $LR_p^{*(1-\alpha)}$) to be used in the test is the $1-\alpha$ quantile of the B bootstrap test statistics, α being the test nominal significance level. The null hypothesis is rejected if $LR_p > LR_p^{*(1-\alpha)}$, where LR_p is the test statistic computed using the original sample. This sampling scheme is known as parametric bootstrap. For simplicity, hereafter, LR_b refers to the bootstrap test.

4 Monte Carlo simulation

In what follows, we report Monte Carlo simulation results on the finite sample behavior of the different maximum likelihood estimators and likelihood ratio tests proposed in this paper. All results were obtained using 10,000 Monte Carlo replications. In each replication, we also performed a bootstrap test based on $B = 1000$ bootstrap samples. Bootstrap sampling was performed parametrically and under the null hypothesis. The sample sizes considered were $n = 25, 35, 45, 55$, the values of the parameter of interest were $\xi = 0.5$, 1.0, 1.5 and the value of σ was fixed at 1.0. All simulations were performed using the `Ox` matrix programming language (Doornik, 2013) and the restricted estimator of σ, $\widehat{\sigma}_\xi$, was obtained using the SQP algorithm, which is implemented in the `MaxSQP` function of the `Ox` programming language.

Table 4 contains simulation results on the following estimators: $\widehat{\xi}_p$, $\widehat{\xi}_{FR}$, $\widehat{\xi}_S$ and $\widehat{\xi}_{CR_a}$. In particular, we report the relative bias (RB), mean squared error (MSE), asymmetry ($\mathcal{A}$) and kurtosis ($\mathcal{K}$) of each estimator. First, notice that the modified profile maximum likelihood estimators ($\widehat{\xi}_{FR}$, $\widehat{\xi}_S$ and $\widehat{\xi}_{CR_a}$) have smaller relative biases than the standard maximum profile likelihood estimator ($\widehat{\xi}_p$). For instance, when $\xi = 1.5$ and $n = 25$ the relative bias of $\widehat{\xi}_p$ is -5.85% whereas those of $\widehat{\xi}_{FR}$, $\widehat{\xi}_S$ and $\widehat{\xi}_{CR_a}$ are -2.35%, -2.27% and -2.70%, respectively. Second, notice that all estimators of ξ are negatively biased. Third, the relative biases decrease when the true value of the shape parameter increases. Fourth, the mean squared errors of all estimators are nearly the same. Finally, as expected, the biases and the mean squared errors decline as the sample size increases.

Table 2 contains the null rejection rates (%) of the profile likelihood ratio (LR_p), modified profile likelihood ratio (LR_{FR}, LR_S and LR_{CR_a}) and bootstrap likelihood ratio (LR_b) tests. The null hypothesis is $\mathcal{H}_0 : \xi = \xi_0$ which is tested against $\mathcal{H}_1 : \xi \neq \xi_0$, where $\xi_0 = 0.5, 1.0, 1.5$. All tests are carried out at four nominal levels, namely: $\alpha = 0.10, 0.05, 0.01, 0.005$. At the outset, we note that the profile likelihood ratio test (LR_p) is considerably oversized, its size distortion (the difference between p-values, estimated by simulation, and nominal levels) decreasing when the value of ξ increases in most scenarios. We also note that the modified profile likelihood ratio tests LR_{FR} and LR_S and the bootstrap test (LR_b) outperform the LR_p test. The LR_{CR_a} test is typically conservative and in some scenarios is more size-distorted than the LR_p test, whereas the LR_{FR}, LR_S tests are slightly liberal and the LR_b test is the least size-distorted. For example, when $n = 25$ and $\xi = 1.0$, the null rejection rates of the LR_p, LR_{FR}, LR_S, LR_{CR_a} and LR_b at the 10% nominal level are, respectively, 12.10%, 10.75%, 10.83%, 8.79% and 10.02%. As expected, the size distortions of all tests decrease when the sample size increases; see Figure 1.

Figure 2 plots the relative quantile discrepancies (i.e., differences between exact and asymptotic quantiles divided by the latter) of the four test statistics against the corresponding asymptotic quantiles. The closer to zero the relative quantile discrepancies, the better the approximation of the exact null distribution by the asymptotic χ^2_1 distribution. The null distributions of the LR_{FR}, LR_S and LR_{CR_a} test statistics are better approximated by the χ^2 asymptotic distribution than that

Table 1 *Point estimation of ξ for $\sigma = 1.0$*

		$\xi = 0.5$				$\xi = 1.0$				$\xi = 1.5$			
n	Estimator	RB	MSE	$\mathcal{A}$	$\mathcal{K}$	RB	MSE	$\mathcal{A}$	$\mathcal{K}$	RB	MSE	$\mathcal{A}$	$\mathcal{K}$
25	$\widehat{\xi}_p$	−0.1895	0.1202	0.0471	3.1149	−0.0879	0.1838	0.1146	3.1309	−0.0585	0.2741	0.2198	3.1066
	$\widehat{\xi}_{FR}$	−0.0896	0.1081	0.1658	3.0798	−0.0399	0.1782	0.1691	3.0992	−0.0235	0.2729	0.2490	3.1004
	$\widehat{\xi}_S$	−0.1139	0.1221	−0.0823	3.4531	−0.0416	0.1839	0.1047	3.1808	−0.0227	0.2776	0.2151	3.1516
	$\widehat{\xi}_{CR_a}$	−0.0856	0.1034	0.2068	3.0173	−0.0427	0.1756	0.1716	3.0506	−0.0270	0.2706	0.2340	3.0521
35	$\widehat{\xi}_p$	−0.1395	0.0822	−0.0053	3.2218	−0.0637	0.1320	0.1200	3.2281	−0.0420	0.1929	0.2043	3.2107
	$\widehat{\xi}_{FR}$	−0.0746	0.0768	0.0758	3.1736	−0.0314	0.1297	0.1544	3.2005	−0.0181	0.1926	0.2200	3.2047
	$\widehat{\xi}_S$	−0.0841	0.0812	−0.0472	3.3324	−0.0319	0.1315	0.1152	3.2522	−0.0177	0.1939	0.2070	3.2138
	$\widehat{\xi}_{CR_a}$	−0.0744	0.0752	0.0862	3.1106	−0.0330	0.1288	0.1493	3.1752	−0.0199	0.1917	0.2072	3.1797
45	$\widehat{\xi}_p$	−0.1031	0.0599	−0.0010	3.1604	−0.0495	0.1010	0.1300	3.1762	−0.0305	0.1485	0.1502	3.1007
	$\widehat{\xi}_{FR}$	−0.0558	0.0572	0.0474	3.1301	−0.0252	0.0999	0.1470	3.1747	−0.0124	0.1486	0.1587	3.0984
	$\widehat{\xi}_S$	−0.0600	0.0589	−0.0118	3.1937	−0.0253	0.1006	0.1324	3.1803	−0.0120	0.1490	0.1509	3.0997
	$\widehat{\xi}_{CR_a}$	−0.0561	0.0567	0.0491	3.0980	−0.0264	0.0995	0.1376	3.1457	−0.0135	0.1482	0.1510	3.0873
55	$\widehat{\xi}_p$	−0.0852	0.0468	0.0384	3.1447	−0.0373	0.0768	0.0999	3.0616	−0.0251	0.1179	0.1351	3.0178
	$\widehat{\xi}_{FR}$	−0.0477	0.0453	0.0685	3.1278	−0.0178	0.0762	0.1102	3.0518	−0.0105	0.1179	0.1401	3.0144
	$\widehat{\xi}_S$	−0.0502	0.0460	0.0361	3.1628	−0.0178	0.0764	0.1021	3.0655	−0.0102	0.1182	0.1366	3.0198
	$\widehat{\xi}_{CR_a}$	−0.0483	0.0450	0.0675	3.1075	−0.0185	0.0761	0.1062	3.0491	−0.0112	0.1177	0.1345	3.0116

Table 2 *Null rejection rates (%), $\sigma = 1.0$ and several samples sizes*

	$\alpha(\%)$	LR_p	LR_{FR}	LR_S	LR_{CR_a}	LR_b	LR_p	LR_{FR}	LR_S	LR_{CR_a}	LR_b
				$n = 25$					$n = 35$		
$\xi = 0.5$	10	13.18	10.55	11.31	7.75	9.66	12.34	10.74	10.94	9.11	10.00
	5	6.98	5.22	5.78	3.01	4.61	6.60	5.41	5.58	4.15	5.12
	1	1.48	0.91	1.08	0.34	0.77	1.67	1.11	1.26	0.68	1.11
	0.5	0.72	0.34	0.50	0.13	0.28	0.90	0.51	0.56	0.23	0.53
$\xi = 1.0$	10	12.10	10.75	10.83	8.79	10.02	12.13	11.03	11.12	9.81	10.77
	5	6.54	5.43	5.69	3.94	5.04	6.56	5.78	5.86	4.90	5.54
	1	1.46	1.01	1.11	0.48	0.91	1.55	1.21	1.28	0.81	1.20
	0.5	0.73	0.49	0.57	0.20	0.46	0.80	0.65	0.64	0.41	0.66
$\xi = 1.5$	10	11.65	10.40	10.42	8.86	9.62	11.23	10.67	10.58	9.60	10.04
	5	5.93	5.14	5.19	4.06	4.72	6.00	5.34	5.35	4.66	5.14
	1	1.35	1.04	1.11	0.61	0.93	1.36	1.25	1.28	0.94	1.19
	0.5	0.67	0.54	0.57	0.27	0.55	0.77	0.65	0.66	0.42	0.68
				$n = 45$					$n = 55$		
$\xi = 0.5$	10	11.66	10.67	10.88	9.38	10.08	11.28	10.42	10.53	9.53	9.98
	5	6.21	5.34	5.45	4.37	5.03	5.79	5.18	5.21	4.54	4.71
	1	1.41	0.97	1.06	0.73	0.97	1.09	0.92	0.99	0.73	0.85
	0.5	0.66	0.56	0.60	0.30	0.54	0.62	0.51	0.54	0.35	0.51
$\xi = 1.0$	10	12.12	11.21	11.31	10.49	10.81	10.97	10.26	10.23	9.65	9.90
	5	6.31	5.60	5.70	5.07	5.55	5.51	5.28	5.18	4.72	4.93
	1	1.50	1.26	1.26	0.95	1.23	1.29	1.12	1.12	0.96	1.13
	0.5	0.70	0.61	0.64	0.53	0.70	0.62	0.55	0.53	0.41	0.54
$\xi = 1.5$	10	11.23	10.83	10.84	10.03	10.42	10.71	10.46	9.76	10.48	10.16
	5	5.74	5.58	5.51	4.94	5.33	5.35	5.25	4.70	5.27	5.00
	1	1.29	1.15	1.14	0.97	1.22	1.25	1.13	0.99	1.12	1.05
	0.5	0.68	0.59	0.58	0.53	0.65	0.65	0.52	0.41	0.52	0.56

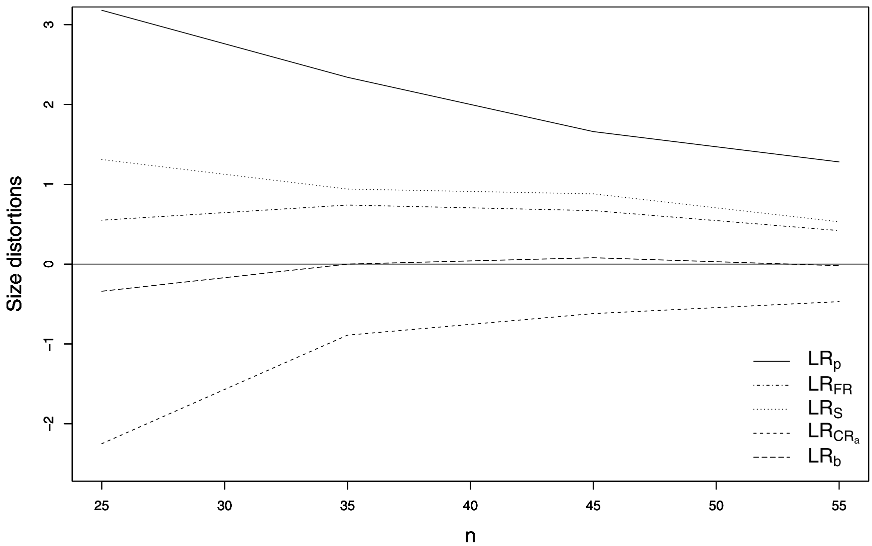

Figure 1 *Size distortions* (%): $\xi = 0.5$, $\alpha = 10\%$.

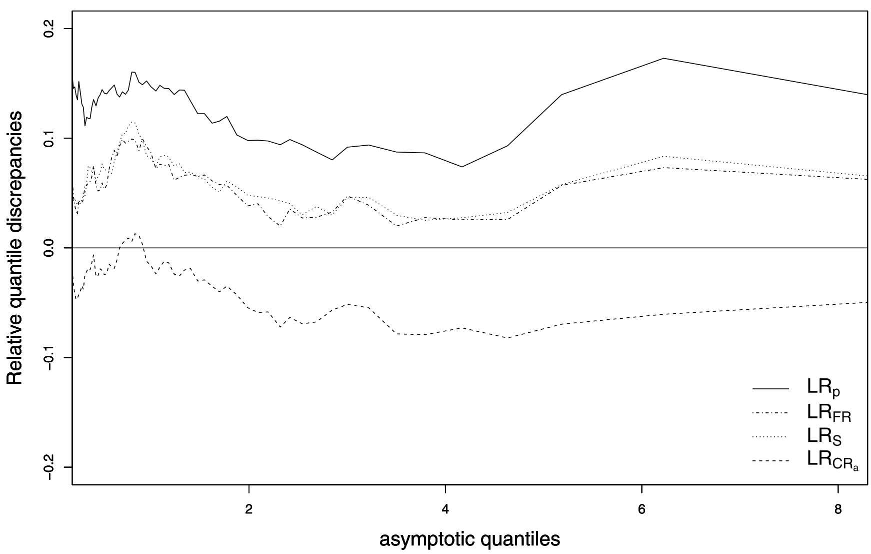

Figure 2 *Relative quantile discrepancy plot*, *inference on* ξ.

of LR_p. The relative discrepancy curves for LR_{FR} and LR_S are quite similar. Table 3 presents the mean and variance of each test statistic when $n = 25$ and $n = 55$ together with the first two χ_1^2 moments. Notice that there is better agreement between exact and asymptotic moments for the modified test statistics LR_{FR} and LR_S relative to LR_p.

Table 3 *Mean and variance of the different test statistics*

ξ		χ_1^2	LR_p	LR_{FR}	LR_S	LR_{CR_a}
				$n = 25$		
0.5	mean	1.00	1.15	1.02	1.06	0.86
	variance	2.00	2.43	1.90	2.22	1.35
1.0	mean	1.00	1.11	1.02	1.03	0.91
	variance	2.00	2.40	2.02	2.07	1.54
1.5	mean	1.00	1.09	1.03	1.03	0.93
	variance	2.00	2.34	2.04	2.07	1.61
				$n = 55$		
0.5	mean	1.00	1.07	1.02	1.02	0.97
	variance	2.00	2.25	2.01	2.04	1.79
1.0	mean	1.00	1.04	1.01	1.01	0.97
	variance	2.00	2.16	2.02	2.02	1.85
1.5	mean	1.00	1.04	1.02	1.02	0.98
	variance	2.00	2.16	2.04	2.04	1.87

We have also performed Monte Carlo simulations under the alternative hypothesis. The empirical nonnull rejection rates (i.e., powers) of the tests are presented in Table 4 at the 10% and 5% nominal levels. Since some of the tests are size-distorted, the tests were performed using size-corrected critical values (obtained from the size simulations) in order to force them to have the correct size. (The bootstrap test is omitted from the analysis since it is not possible to size-correct it.) The power comparisons were performed for several values of ξ, $\sigma = 1.0$ and $n = 25$. The figures in Table 4 show that the LR_p test is the least powerful and that the LR_{CR_a} test is slightly more powerful than the other modified tests. The power curves are displayed in Figure 3. As expected, the tests become more powerful as the true parameter value moves away from the value specified in the null hypothesis.

5 Illustration

We shall now perform profile and modified profile likelihood inference using two real data sets. The `Ox` code for the data analysis is available at http://www.de.ufpb.br/~juliana/pareto.html.

For the first illustration we consider data on annual floods of the Nidd River at Hunsingore, England, from 1934 to 1969 (35 years). These data were analyzed by Hosking and Wallis (1987). The authors fitted the GPD to the excesses of the Nidd peak floods over different thresholds using the probability weighted moments method. We present in Table 5 the profile and modified profile maximum likelihood estimates of ξ for the following thresholds: $t = 70, 80, 90, 100$. Except for $t = 100$, all point estimates are quite similar for each of the other threshold values.

Table 4 *Nonnull rejection rates* (%), *inference on* ξ

		$\alpha = 10\%$				$\alpha = 5\%$			
ξ_0	ξ	LR_p	LR_{FR}	LR_S	LR_{CR_a}	LR_p	LR_{FR}	LR_S	LR_{CR_a}
0.5	0.6	10.34	11.55	11.41	12.15	5.18	6.65	6.52	7.25
	0.8	19.68	23.80	23.37	25.47	12.63	16.15	18.75	16.86
	1.0	35.78	41.54	40.85	43.57	26.17	31.93	35.84	32.67
	1.2	53.17	58.83	58.01	60.78	43.46	49.55	53.75	50.74
	1.4	67.60	72.42	71.36	73.71	59.27	64.53	67.78	65.74
	1.6	79.27	83.13	82.48	83.98	72.60	76.58	79.48	77.62
	1.8	87.04	89.72	89.11	90.33	82.06	85.33	87.16	86.20
	2.0	91.50	93.24	92.82	93.67	88.01	90.17	91.55	90.78
	3.0	99.13	99.44	99.35	99.46	98.61	98.97	99.10	99.10
1.0	1.2	12.30	14.01	14.10	14.56	6.56	8.07	8.14	8.69
	1.4	20.44	23.65	23.85	24.51	12.76	15.76	15.94	16.91
	1.6	32.31	36.42	36.67	37.66	22.69	27.16	27.42	28.76
	1.8	45.84	50.48	50.62	51.52	35.14	40.61	40.87	42.38
	2.0	58.75	63.16	63.12	63.94	48.43	53.55	53.63	55.16
	2.2	70.20	73.86	73.82	74.31	60.86	65.71	65.88	67.22
	2.6	85.27	87.38	87.46	87.65	79.28	82.57	82.40	83.11
	3.0	92.73	93.93	93.87	94.08	89.45	91.47	91.40	91.69
	4.0	99.14	99.32	99.26	99.31	98.32	98.80	98.73	98.82
1.5	1.8	13.64	15.81	15.92	16.09	8.11	9.56	9.74	9.91
	2.0	21.31	24.39	24.29	24.57	13.44	16.08	16.50	16.78
	2.2	30.80	34.93	34.80	35.43	21.76	25.07	25.35	26.01
	2.4	42.14	46.29	46.38	46.66	31.45	35.85	36.33	36.88
	2.6	53.57	57.59	57.66	57.91	42.63	47.01	47.38	48.09
	2.8	63.61	67.19	67.31	67.81	53.64	57.89	58.27	59.00
	3.0	71.18	74.45	74.41	74.63	62.42	66.24	66.51	67.05
	4.0	93.76	94.79	94.72	94.74	90.68	92.13	92.14	92.37
	5.0	98.89	99.11	99.08	99.12	98.04	98.47	98.47	98.51

We also note that the estimates based on the modified profile likelihood functions are slightly larger than the profile maximum likelihood estimate. Figure 4 shows the profile and modified profile log-likelihood functions as functions of ξ and the threshold $t = 80$.

Suppose we are interested in testing $\mathcal{H}_0 : \xi = 0.11$ against $\mathcal{H}_1 : \xi \neq 0.11$. Table 6 contains the p-values of the LR_p, LR_{FR}, LR_S and LR_{CR_a} tests. Notice that for $t = 80$ the conclusion is reversed when based on tests derived from the modified profile log-likelihood functions: the null hypothesis is not rejected at the 10% nominal level by the profile likelihood ratio test but it is rejected by the modified profile likelihood ratio tests. It is noteworthy that one can compute and analyze return levels; for details, see Coles (2001).

In the second illustration we use financial data provided by Coles (2001). The data are on daily closing prices of the Dow Jones Index from 1996 to

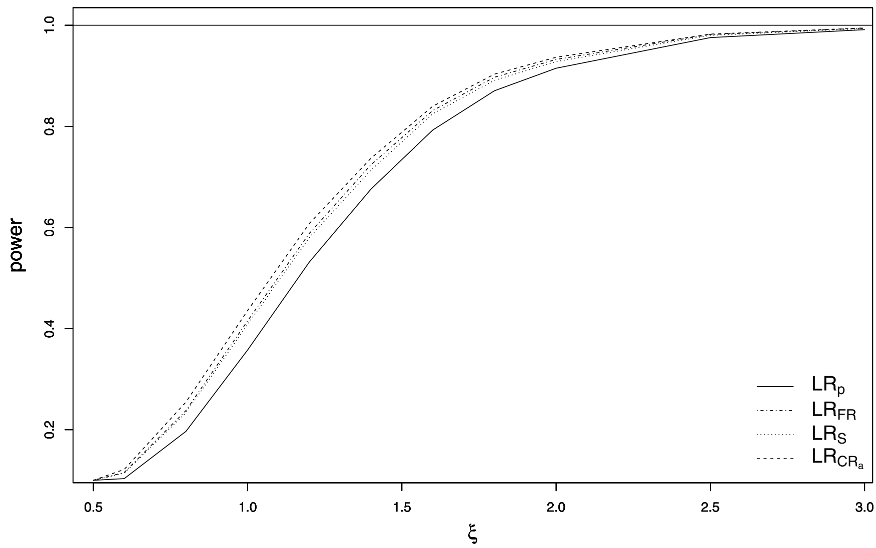

Figure 3 *Powers of the tests*: $\xi = 0.5$, $\alpha = 10\%$ *and* $n = 25$.

Table 5 *Profile and modified profile maximum likelihood estimates of* ξ

t	$\widehat{\xi}_p$	$\widehat{\xi}_{FR}$	$\widehat{\xi}_S$	$\widehat{\xi}_{CR_a}$
100	0.0033	0.0471	0.0204	0.0504
90	0.2383	0.2649	0.2527	0.2630
80	0.3429	0.3568	0.3543	0.3556
70	0.3232	0.3295	0.3312	0.3296

2000. Since the series is non-stationary, he suggests transforming the series as $\widetilde{y}_i = \log(y_i) - \log(y_{i-1})$. He also suggests setting the threshold at $t = 2$ which leads to 37 exceedances out of 1303 data points. The profile and modified profile maximum likelihood estimates of ξ are $\widehat{\xi}_p = 0.2878$, $\widehat{\xi}_{FR} = 0.3271$, $\widehat{\xi}_S = 0.3124$ and $\widehat{\xi}_{CR_a} = 0.3243$. In accordance to our simulation results, the maximum likelihood estimator underestimates ξ. Suppose we wish to test $\mathcal{H}_0 : \xi = 0.85$ against $\mathcal{H}_1 : \xi \neq 0.85$. The test statistics are $LR_p = 2.941$, $LR_{FR} = 2.476$, $LR_S = 2.594$ and $LR_{CR_a} = 2.304$, and the respective p-values are $0.086, 0.116, 0.107$ and 0.129. Therefore, the profile likelihood ratio test rejects the null hypothesis at the 10% nominal level, unlike the modified profile likelihood ratio tests. Thus, the unmodified and modified tests yield different inferences.

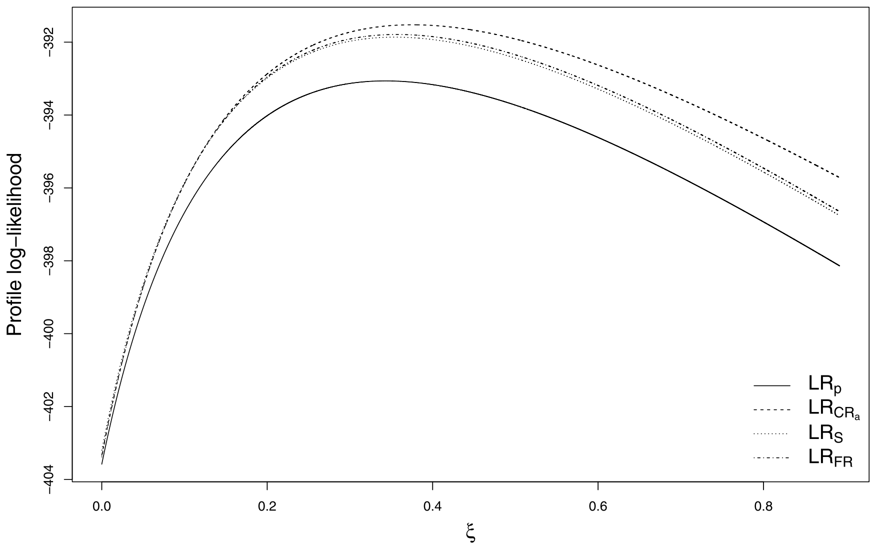

Figure 4 *Profile and modified profile log-likelihood functions for ξ.*

Table 6 *p-values*

t	LR_p	LR_{FR}	LR_S	LR_{CR_a}
100	0.638	0.781	0.703	0.800
90	0.499	0.415	0.462	0.432
80	0.102	0.084	0.091	0.090
70	0.029	0.025	0.025	0.026

6 Concluding remarks

We considered the issue of performing inference on the GPD shape parameter. We obtained adjustments to the profile likelihood function. Inferences based on such functions are expected to be considerably more accurate than those based on the standard profile likelihood function. Bootstrap inference was also considered. We presented and discussed the results of Monte Carlo simulations on point estimation and hypothesis testing inference. The numerical evidence favors the estimators and tests based on the modified profile likelihood functions and also the bootstrap test. The profile maximum likelihood estimator can be severely biased and the profile likelihood ratio test can be quite liberal in small samples. The modified estimators proposed in this paper displayed smaller biases and the modified tests were less size-distorted. We therefore recommend the modified profile likelihood inference proposed in the present paper be used in empirical applications that employ the GPD. A direction for future research is the development of Bayesian inferential procedures for the GPD. Another direction for future research is the development of testing inference based on the signed likelihood ratio statistic, including higher order modifications.

Acknowledgments

We thank three anonymous referees for their comments and suggestions. We also gratefully acknowledge financial support from CAPES and CNPq.

References

Barndorff-Nielsen, O. (1983). On a formula to the distribution of the maximum likelihood estimator. *Biometrika* **70**, 343–365. MR0712023

Castillo, E. and Hadi, A. S. (1997). Fitting the generalized Pareto distribution to data. *Journal of the American Statistical Association* **92**, 1609–1620. MR1615270

Castillo, E., Hadi, A. S., Balakrishnan, N. and Sarabia, J. M. (2005). *Extreme Value and Related Models with Applications in Engineering and Science*. Hoboken, NJ: Wiley. MR2191401

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: Springer. MR1932132

Cox, D. and Hinkley, D. (1974). *Theoretical Statistics*. London: Chapman & Hall. MR0370837

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society. Series B* **49**, 1–39. MR0893334

Cox, D. R. and Reid, N. (1993). A note on the calculation of adjusted profile likelihood. *Journal of the Royal Statistical Society. Series B* **55**, 467–471. MR1224410

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society. Series B* **30**, 248–275. MR0237052

Davison, A. C. (2003). *Statistical Models*. Cambridge: Cambridge Univ. Press. MR1998913

de Carvalho, M., Turkman, K. F. and Rua, A. (2013). Dynamic threshold modelling and the US business cycle. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **62**, 535–550. MR3083910

Doornik, J. A. (2013). *An Object-Oriented Matrix Language Ox 7*. London: Timberlake Consultants. Available at http://www.doornik.com/.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* **7**, 1–26. MR0515681

Ferreira, A. and Haan, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli* **20**, 1717–1737. MR3263087

Fraser, D. A. S. and Reid, N. (1995). Ancillaries and third-order significance. *Utilitas Mathematica* **47**, 33–53. MR1330888

Fraser, D. A. S., Reid, N. and Wu, J. (1999). A simple general formula for tail probabilities for frequentist and Bayesian inference. *Biometrika* **86**, 249–264. MR1705367

Giles, D. E., Feng, H. and Godwin, R. T. (2011). Bias-corrected maximum likelihood estimation of the parameters of the generalized Pareto distribution. Econometrics Working Papers 1105, Department of Economics, University of Victoria.

Hosking, J. R. M. and Wallis, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* **29**, 339–349. MR0906643

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. New York: Springer. MR2244940

Pace, L. and Salvan, A. (1997). *Principles of Statistical Inference from a Neo-Fisherian Perspective*. Singapore: World Scientific. MR1476674

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics* **3**, 119–131. MR0423667

Severini, T. A. (1999). An empirical adjustment to the likelihood ratio statistic. *Biometrika* **86**, 235–247. MR1705371

Severini, T. A. (2000). *Likelihood Methods in Statistics*. Oxford: Oxford University Press. MR1854870

J. F. Pires
Departamento de Estatística
Universidade Federal da Paraíba
João Pessoa-PB, CEP 58051-900
Brasil
E-mail: juliana@de.ufpb.br

A. H. M. A. Cysneiros
F. Cribari-Neto
Departamento de Estatística
Universidade Federal de Pernambuco
Recife-PE, CEP 50740-540
Brasil
E-mail: audrey@de.ufpe.br
cribari@de.ufpe.br