University of Victoria
Department of Economics

Econometrics Working Paper EWP1105

ISSN 1485-6441

Bias - Corrected Maximum Likelihood Estimation of the Parameters of the Generalized Pareto Distribution

David E. Giles
Department of Economics, University of Victoria

Hui Feng
Department of Economics, Business & Mathematics
King's University College
University of Western Ontario
&

Ryan T. Godwin
Department of Economics, University of Victoria

Revised, October 2011

Abstract

We derive analytic expressions for the biases, to $O(n^{-1})$, of the maximum likelihood estimators of the parameters of the generalized Pareto distribution. Using these expressions to bias-correct the estimators in a selective manner is found to be extremely effective in terms of bias reduction, and can also result in a small reduction in relative mean squared error. In terms of remaining relative bias, the analytic bias-corrected estimators are somewhat less effective than their counterparts obtained by using a parametric bootstrap bias correction. However, the analytic correction out-performs the bootstrap correction in terms of remaining %MSE. It also performs credibly relative to other recently proposed estimators for this distribution. Taking into account the relative computational costs, this leads us to recommend the selective use of the analytic bias adjustment for most practical situations.

Keywords: Maximum likelihood; Bias reduction; Extreme values; Generalized Pareto distribution; Peaks over threshold; Parametric bootstrap

Mathematics Subject Classification: 62F10; 62F40; 62N02; 62N05

Author Contact:
David E. Giles, Dept. of Economics, University of Victoria, P.O. Box 1700, STN CSC, Victoria, B.C., Canada

1. Introduction

This paper discusses the calculation of analytic second-order bias expressions for the maximum likelihood estimators (MLEs) of the parameters of the generalized Pareto distribution (GPD). This distribution is widely used in extreme value analysis in many areas of application. These include empirical finance (*e.g.*, Angelini, 2002; Klüppelberg, 2002; Bali and Neftci, 2003; Gilli and Këllizi, 2006; and Gençay and Selçuk, 2006; Ren and Giles, 2010); meteorology (*e.g*., Holmes and Moriarty, 1999); hydrology (*e.g*., Van Montfort and Witter, 1986); climatology (*e.g*., Nadarajah, 2008); metallurgy (*e.g*., Shi *et al*., 1999); seismology (*e.g*., Pisarenko and Sornette, 2003; and Huyes *et al*., 2010); actuarial science (*e.g*., Cebriàn *et al*., 2003; Brazouskas and Kleefeld, 2009); ocean science (*e.g*., Stansell, 2005); and movie box office revenues (Bi and Giles, 2009). A useful summary table of additional applications is provided by de Zea Bermudez and Kotz (2010, p.1370).

The motivation for the use of the GPD in such studies arises from asymptotic theory that is specific to the tail behaviour of the data. Accordingly, in practice, the parameters may be estimated from a relatively small number of extreme order statistics (as is the case if the so-called "peaks over threshold" procedure is used), so the finite-sample properties of the MLEs for the parameters of this distribution are of particular interest. Some attention has been paid previously to the small-sample bias of the MLEs for this distribution, most notably by Hosking and Wallis (1987). However, the earlier evidence is entirely simulation-based, and in many cases must be viewed with caution because of subsequently recognized issues associated with the maximization of the likelihood function. In this paper we consider the $O(n^{-1})$ bias formula introduced by Cox and Snell (1968). This methodology is especially appealing here, as it enables us to obtain analytic bias expressions, and hence "bias-corrected" MLEs, even though the likelihood equations for the GPD *do not* admit a closed-form solution.

It should be noted that the Cox-Snell approach that we adopt here is "corrective", in the sense that a "bias adjusted" MLE can be constructed by subtracting the bias (estimated at the MLEs of the parameters) from the original MLE. An alternative "preventive" approach, introduced by Firth (1993), involves modifying the score vector of the log-likelihood function *prior* to solving for the MLE. Interestingly, Cribari-Neto and Vasconcellos (2002) find that these two approaches are equally successful with respect to (finite sample) bias reduction without loss of efficiency in the context of the MLE for the parameters of the beta distribution. In that same context, they find that the bootstrap performs poorly with respect to bias reduction and efficiency. We do not pursue preventive methods of bias reduction in this study.

Our results show that bias-correcting the MLEs for the parameters of the GPD, using the estimated values of the analytic $O(n^{-1})$ bias expressions, is extremely effective in reducing absolute relative bias. In addition, this is often accompanied by a modest reduction in relative mean squared error. We compare this analytic bias correction with the alternative of using the parametric bootstrap to estimate the $O(n^{-1})$ bias, and then correcting accordingly. We find that the bootstrap bias-correction can be even more effective in terms of reducing bias, but this generally comes at the expense of increased relative mean squared error. Also, in practice its application raises some computational issues. Consequently we do not recommend a bootstrap bias correction for the MLEs of the GPD parameters, and instead favour the Cox-Snell analytic correction.

Section 2 summarizes the required background theory, which is then used to derive analytic expressions for the first-order biases of the MLEs of the parameters of the generalized Pareto distribution in section 3. Section 4 reports the results of a simulation experiment that evaluates the properties of bias-corrected estimators that are based on our analytic results, as well as the corresponding bootstrap bias-corrected MLEs. Some illustrative applications are provided in section 5, and some concluding remarks appear in section 6.

2. Second-order biases of maximum likelihood estimators

For some arbitrary distribution, let $l(\theta)$ be the (total) log-likelihood based on a sample of n observations, with p-dimensional parameter vector, θ. $l(\theta)$ is assumed to be regular (in the sense of Dugué, 1937 and Cramér, 1946, p.500) with respect to all derivatives up to and including the third order. The implications of this in the case of the GPD are discussed below.

The joint cumulants of the derivatives of $l(\theta)$ are denoted:

$$k_{ij} = E(\partial^2 l / \partial\theta_i \partial\theta_j) \quad ; \quad i, j = 1, 2, \ldots, p \tag{1}$$

$$k_{ijl} = E(\partial^3 l / \partial\theta_i \partial\theta_j \partial\theta_l) \quad ; \quad i, j, l = 1, 2, \ldots, p \tag{2}$$

$$k_{ij,l} = E[(\partial^2 l / \partial\theta_i \partial\theta_j)(\partial l / \partial\theta_l)] \quad ; \quad i, j, l = 1, 2, \ldots, p \tag{3}$$

and the derivatives of the cumulants are:

$$k_{ij}^{(l)} = \partial k_{ij} / \partial\theta_l \quad ; \quad i, j, l = 1, 2, \ldots, p. \tag{4}$$

All of the ‘k’ expressions are assumed to be $O(n)$.

Extending earlier work by Bartlett (1953a, 1953b), Haldane (1953), Haldane and Smith (1956), Shenton and Wallington (1962) and Shenton and Bowman (1963), Cox and Snell (1968) showed that when the sample data are independent (but not necessarily identically distributed) the bias of the s^{th} element of the MLE of θ ($\hat{\theta}$) is:

$$Bias(\hat{\theta}_s) = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{l=1}^{p} k^{si}k^{jl}[0.5k_{ijl} + k_{ij,l}] + O(n^{-2})\ ; \ \ s = 1, 2, \ldots, p. \qquad (5)$$

where k^{ij} is the $(i,j)^{th}$ element of the inverse of the (expected) information matrix, $K = \{-k_{ij}\}$. Cordeiro and Klein (1994) noted that this bias expression also holds if the data are non-independent, provided that all of the k terms are $O(n)$, and that it can be re-written as:

$$Bias(\hat{\theta}_s) = \sum_{i=1}^{p} k^{si} \sum_{j=1}^{p}\sum_{l=1}^{p} [k_{ij}^{(l)} - 0.5k_{ijl}]k^{jl} + O(n^{-2})\ ;\ s = 1, 2, \ldots, p. \qquad (6)$$

Notice that (6) has a computational advantage over (5), as it does not involve terms of the form defined in (3).

Now, let $a_{ij}^{(l)} = k_{ij}^{(l)} - (k_{ijl}/2)$, for $i, j, l = 1, 2, \ldots, p$; and define the following matrices:

$$A^{(l)} = \{a_{ij}^{(l)}\}\ ; \quad i, j, l = 1, 2, \ldots, p \qquad (7)$$

$$A = [A^{(1)} \mid A^{(2)} \mid \ldots\ldots \mid A^{(p)}]. \qquad (8)$$

Cordeiro and Klein (1994) show that the expression for the $O(n^{-1})$ bias of the MLE of θ ($\hat{\theta}$) can be re-written in the convenient matrix form:

$$Bias(\hat{\theta}) = K^{-1}A\,vec(K^{-1}) + O(n^{-2}). \qquad (9)$$

A “bias-corrected” MLE for θ can then be obtained as:

$$\tilde{\theta} = \hat{\theta} - \hat{K}^{-1}\hat{A}\ vec(\hat{K}^{-1}), \qquad (10)$$

where $\hat{K} = (K)|_{\hat{\theta}}$ and $\hat{A} = (A)|_{\hat{\theta}}$, and the bias of $\tilde{\theta}$ is $O(n^{-2})$.

3. Bias correction for the generalized Pareto distribution

We now turn to the problem of reducing the bias of the MLEs for the parameters of a distribution that is widely used in the context of the peaks-over-threshold method in extreme value analysis. The generalized Pareto distribution (GPD) was proposed by Pickands (1975), and it follows directly from the generalized extreme value (GEV) distribution (Coles, 2001, pp.47-48, 75-76) that is used in the context of block maxima data. The distribution and density functions for the GPD, with shape parameter, or tail index, ξ and scale parameter σ, are:

$$
\begin{aligned}
F(y) &= 1-\left(1+\xi y/\sigma\right)^{-1/\xi}; \quad y>0,\ \xi \neq 0 \\
&= 1-\exp(-y/\sigma)\,; \qquad\qquad \xi = 0
\end{aligned}
\tag{11}
$$

$$
\begin{aligned}
f(y) &= (1/\sigma)\left(1+\xi y/\sigma\right)^{-1/\xi-1}; \quad y>0,\ \xi \neq 0 \\
&= (1/\sigma)\exp(-y/\sigma)\,; \qquad\qquad \xi = 0
\end{aligned}
\tag{12}
$$

respectively. Note that $0 \le y < \infty$ if $\xi \ge 0$, and $0 \le y < -\sigma/\xi$ if $\xi < 0$. The (integer-order) central moments of the GPD can be shown (*e.g.*, Arnold, 1983, pp. 50-51) to be:

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

and the r^{th} moment exists if $\xi < 1/r$.

We will be concerned with the MLE for $\theta' = (\xi, \sigma)$. The finite-sample properties of this estimator have not been considered *analytically* before, although Hosking and Wallis (1987) and Moharram *et al*. (1993) provide some simulation evidence. Jondeau and Rockinger (2003) provide some limited Monte Carlo results for a modified MLE of the shape parameter in the related GEV distribution. Other estimators are available, and many of them are reviewed by de Zea Bermudez and Kotz (2010). Hosking and Wallis (1987) discuss the method of moments (MOM) and probability-weighted moments (PWM) estimators of θ; Castillo and Hadi (1997) propose the “elemental percentile method”; Luceño (2006) considers various “maximum goodness of fit” estimators, based on the empirical distribution function; and Brazauskas and Kleefeld (2009) provide a robust estimation procedure. Finally, Zhang (2007) proposes a “likelihood moment estimator”, and Zhang and Stephens (2009) discuss a quasi-Bayesian estimator. We discuss these last two estimators further in section 4, as we compare their performance with that of our bias corrected estimator in this study.

The above condition for the existence of moments can, of course, limit the applicability of MOM estimation for this distribution. In what follows, it is important to note that the MLE is also defined only in certain parameter ranges. More specifically, the MLEs of ξ and σ do not exist if $\xi < -1$ because in that case the density in (4.2) tends to infinity when y tends to $-\sigma / \xi$. In addition, the usual regularity conditions do not hold if $\xi < -1/2$ (Smith, 1985, p.89). For these and other reasons, maximum likelihood estimation of the parameters of the GPD can be challenging in practice, as is discussed more fully by Davison and Smith (1990), Grimshaw (1993), Castillo and Hadi (1997), Chaouche and Bacro (2006), Castillo and Daoudi (2009), and Zhang and Stephens (2009).

Assuming independent observations, which in practice may require that the data be "de-clustered" prior to use, the full log-likelihood based on (12) is:

$$l(\xi,\sigma) = -n\ln(\sigma) - (1+1/\xi)\sum_{i=1}^{n}\ln(1+\xi y_i/\sigma) \; . \qquad (13)$$

So,

$$\partial l/\partial\xi = \xi^{-2}\sum_{i=1}^{n}\ln(1+\xi y_i/\sigma) - (1+\xi^{-1})\sum_{i=1}^{n}[y_i/(\sigma+\xi y_i)] \qquad (14)$$

$$\partial l/\partial\sigma = \sigma^{-1}\{-n + (1+\xi)\sum_{i=1}^{n}[y_i/(\sigma+\xi y_i)]\} \qquad (15)$$

$$\partial^2 l/\partial\xi^2 = 2\xi^{-3}\{\xi\sum_{i=1}^{n}y_i/(\sigma+\xi y_i) - \sum_{i=1}^{n}[\ln(1+\xi y_i/\sigma)]\} + (1+\xi^{-1})\sum_{i=1}^{n}[y_i^2/(\sigma+\xi y_i)^2] \qquad (16)$$

$$\partial^2 l/\partial\sigma^2 = \sigma^{-2}\{n - (1+\xi)\sum_{i=1}^{n}[y_i(2\sigma+\xi y_i)/(\sigma+\xi y_i)^2]\} \qquad (17)$$

$$\partial^2 l/\partial\xi\partial\sigma = \xi^{-1}\{(1+\xi)\sum_{i=1}^{n}[y_i/(\sigma+\xi y_i)^2] - \sigma^{-1}\sum_{i=1}^{n}[y_i/(\sigma+\xi y_i)]\} \qquad (18)$$

$$\begin{aligned}\partial^3 l/\partial\xi^3 = 3\xi^{-4}\{&2\sum_{i=1}^{n}[\ln(1+\xi y_i/\sigma)] - 2\xi\sum_{i=1}^{n}[y_i/(\sigma+\xi y_i)] \\ &- \xi^2\sum_{i=1}^{n}[y_i^{\;2}/(\sigma+\xi y_i)^2]\} - 2(1+1/\xi)\sum_{i=1}^{n}[(y_i^3/(\sigma+\xi y_i))^3]\end{aligned} \qquad (19)$$

$$\begin{aligned}\partial^3 l/\partial\sigma^3 = 2\sigma^{-3}\{&-n + (1+\xi)\sum_{i=1}^{n}[y_i(2\sigma+\xi y_i)/(\sigma+\xi y_i)^2]\} \\ &+ 2\sigma^{-1}(1+\xi)[\sum_{i=1}^{n}[y_i/(\sigma+\xi y_i)^3]\end{aligned} \qquad (20)$$

$$\begin{aligned}\partial^3 l/\partial\xi^2\partial\sigma = -2\xi^{-2}\{&\sum_{i=1}^{n}[y_i/(\sigma+\xi y_i)^2] - \sigma^{-1}\sum_{i=1}^{n}[y_i/(\sigma+\xi y_i)]\} \\ &- 2(1+1/\xi)\sum_{i=i}^{n}[y_i^{\;2}/(\sigma+\xi y_i)^3]\end{aligned} \qquad (21)$$

$$\partial^3 l / \partial\xi\partial\sigma^2 = \xi^{-1}\{\sigma^{-2}\sum_{i=1}^{n}[y_i/(\sigma+\xi y_i)] + \sigma^{-1}\sum_{i=1}^{n}[y_i/(\sigma+\xi y_i)^2] - 2(1+\xi)\sum_{i=1}^{n}[y_i/(\sigma+\xi y_i)^3]\} \quad (22)$$

The first-order conditions that are obtained by setting (14) and (15) to zero do not admit a closed-form solution. However, we can still determine the bias of the MLEs of the parameters and then obtain their "bias-adjusted" counterparts by modifying the numerical solutions (estimates) to the likelihood equations by the extent of the (estimated) bias.

The following results are obtained readily by direct integration after the change of variable, $x = 1 + \xi y / \sigma$, regardless of the sign of ξ, and hence regardless of the domains of *y* and *x*:

$$E[y/(\sigma+\xi y)] = (1+\xi)^{-1} \quad ; \quad \xi > -1 \quad (23)$$

$$E[y/(\sigma+\xi y)^2] = [\sigma(1+\xi)(1+2\xi)]^{-1} \quad ; \quad \xi > -1/2 \quad (24)$$

$$E[y/(\sigma+\xi y)^3] = [\sigma^2(1+2\xi)(1+3\xi)]^{-1} \quad ; \quad \xi > -1/3 \quad (25)$$

$$E[y^2/(\sigma+\xi y)^2] = 2[(1+\xi)(1+2\xi)]^{-1} \quad ; \quad \xi > -1/2 \quad (26)$$

$$E[y^2/(\sigma+\xi y)^3] = 2[\sigma(1+\xi)(1+2\xi)(1+3\xi)]^{-1} \quad ; \quad \xi > -1/3 \quad (27)$$

$$E[y^3/(\sigma+\xi y)^3] = 6[(1+\xi)(1+2\xi)(1+3\xi)]^{-1} \quad ; \quad \xi > -1/3 \quad (28)$$

$$E[y(2\sigma+\xi y)/(\sigma+\xi y)] = 2\sigma[(1-\xi)(1+\xi)]^{-1} \quad ; \quad -1 < \xi < 1 \quad (29)$$

$$E[y(2\sigma+\xi y)/(\sigma+\xi y)^2] = 2[(1+2\xi)]^{-1} \quad ; \quad \xi > -1/2 \quad (30)$$

$$E[y(2\sigma+\xi y)/(\sigma+\xi y)^3] = 2[\sigma(1+\xi)(1+3\xi)]^{-1} \quad ; \quad \xi > -1/3 \quad . \quad (31)$$

Recalling that $0 \le y < -\sigma/\xi$ if $\xi < 0$, the constraints associated with ξ in equations (23) to (31) ensure the existence and positivity of the various expectations. Collectively, these constraints require that $-1/3 < \xi < 1$. To ensure, in addition, the existence of the first two (three) moments of *y* we require that $-1/3 < \xi < 1/2$ ($-1/3 < \xi < 1/3$). With the change of variable, $z = \ln(1 + \xi y / \sigma)$, and using formula 3.381 no. 4 from Gradshteyn and Ryzhik (1965, p.317) with $\xi > 0$, we also have:

$$E[\ln(1+\xi y/\sigma)] = \xi \quad . \quad (32)$$

It is readily shown that (32) also holds for $\xi < 0$.

We can now evaluate the various terms needed to determine the Cox-Snell biases of the MLEs of ξ and σ, as discussed in section 2:

$$k_{11} = -2n/[(1+\xi)(1+2\xi)] \tag{33}$$

$$k_{22} = -n/[\sigma^2(1+2\xi)] \tag{34}$$

$$k_{12} = -n/[\sigma(1+\xi)(1+2\xi)] \tag{35}$$

$$k_{111} = 24n/[(1+\xi)(1+2\xi)(1+3\xi)] \tag{36}$$

$$k_{222} = 4n/[\sigma^3(1+3\xi)] \tag{37}$$

$$k_{112} = 8n/[\sigma^2(1+\xi)(1+2\xi)(1+3\xi)] \tag{38}$$

$$k_{122} = 4n/[\sigma^2(1+2\xi)(1+3\xi)] \tag{39}$$

$$k_{11}^{(1)} = 2n(3+4\xi)/[(1+\xi)^2(1+2\xi)^2] \tag{40}$$

$$k_{11}^{(2)} = 0 \tag{41}$$

$$k_{22}^{(1)} = 2n/[\sigma^2(1+2\xi)^2] \tag{42}$$

$$k_{22}^{(2)} = 2n/[\sigma^3(1+2\xi)] \tag{43}$$

$$k_{12}^{(1)} = n(3+4\xi)/[\sigma(1+\xi)^2(1+2\xi)^2] \tag{44}$$

$$k_{12}^{(2)} = n/[\sigma^2(1+\xi)(1+2\xi)] \tag{45}$$

Note that all of (33) to (45) are $O(n)$, as is required for the Cox-Snell result. The information matrix is

$$K = n\begin{bmatrix} 2/[(1+\xi)(1+2\xi)] & 1/[\sigma(1+\xi)(1+2\xi)] \\ 1/[\sigma(1+\xi)(1+2\xi)] & 1/[\sigma^2(1+2\xi)] \end{bmatrix} . \tag{46}$$

The elements of $A^{(1)}$ are:

$$a_{11}^{(1)} = 2n(3+4\xi)/[(1+\xi)^2(1+2\xi)^2] - 12n/[(1+\xi)(1+2\xi)(1+3\xi)] \tag{47}$$

$$a_{22}^{(1)} = 2n/[\sigma^2(1+2\xi)^2] - 2n/[\sigma^2(1+2\xi)(1+3\xi)] \tag{48}$$

$$a_{12}^{(1)} = a_{21}^{(1)} = n(3+4\xi)/[\sigma(1+\xi)^2(1+2\xi)^2] - 4n/[\sigma^2(1+\xi)(1+2\xi)(1+3\xi)] \quad , \tag{49}$$

and the corresponding elements of $A^{(2)}$ are:

$$a_{11}^{(2)} = -4n/[\sigma^2(1+\xi)(1+2\xi)(1+3\xi)] \tag{50}$$

$$a_{22}^{(2)} = 2n/[\sigma^3(1+2\xi)] - 2n/[\sigma^3(1+3\xi)] \tag{51}$$

$$a_{12}^{(2)} = a_{21}^{(2)} = n/[\sigma^2(1+\xi)(1+2\xi)] - 2n/[\sigma^2(1+2\xi)(1+3\xi)] \quad (52)$$

Defining $A = [A^{(1)} \mid A^{(2)}]$, the expression for the biases of the MLEs of ξ and σ to order $O(n^{-1})$ is

$$B = Bias\begin{pmatrix}\hat{\xi} \\ \hat{\sigma}\end{pmatrix} = K^{-1}A\,vec(K^{-1})\,, \quad (53)$$

which can be evaluated by using (46) to (52), provided that $-1/3 < \xi < 1$.

Noting that all of the $a_{ij}^{(l)}$ terms are of order *n*, and that (from (46)) K^{-1} is of order n^{-1}, we see that the bias expression in (53) is indeed $O(n^{-1})$, as required. Finally, a "bias-corrected" MLE for the parameter vector can be obtained as $(\tilde{\xi}, \tilde{\sigma})' = (\hat{\xi}, \hat{\sigma})' - \hat{B}'$, where $\hat{B}$ is constructed by replacing ξ and σ in (53) with their MLEs. This modified estimator is unbiased to order $O(n^{-2})$, but should not be used unless $-1/3 < \hat{\xi} < 1$.

4. Simulation results

The bias expression in (53) is valid only to $O(n^{-1})$. The actual bias and mean squared error (MSE) of the maximum likelihood and bias-corrected maximum likelihood estimators have been evaluated in a Monte Carlo experiment. The simulations were undertaken using the *R* statistical software environment (R, 2008). Generalized Pareto variates were generated using the *evd* package (Stephenson, 2008), and the log-likelihood function was maximized using the method outlined by Grimshaw (1993) using *R* (2008) code kindly supplied by the latter author. Grimshaw's algorithm was used as it deals carefully with several known difficulties that arise with MLE in the context of the GPD. Earlier simulation experiments that ignore the subtleties associated with this MLE problem should be viewed with caution. Each part of our experiment uses 50,000 Monte Carlo replications.

Without loss of generality, we have set $\sigma = 1$. The sample sizes that we consider are motivated by practical applications. For example, Brooks *et al*. (2005) and Brazouskas and Kleefled (2009) deal with (effective) sample sizes as small as *n* = 35, 40, Nadarajah (2008) uses samples ranging from 66 to 90, while Bali and Neftci (2003) have a sample of *n* = 300. We report results for several values of ξ that are consistent with the validity of our bias correction formula, and with the existence of the first two moments of the GPD. Positive values of ξ are especially pertinent in the modeling of returns on financial assets (*e.g*., see the results of Klüppelberg, 2002; Bali and Neftci, 2003; Gilli and Këllizi, 2006; and Gençay and Selçuk, 2006), and extremes in wind gusts (Holmes and Moriarty, 1999). In practice, positive values of $\hat{\xi}$ pose no special computational issues when bias-correcting the MLEs of the parameters. Negative values for the shape parameter are also considered, as they arise in other areas of application such as hydrology (Van Montfort and Witter, 1986), climatology (Nadarajah, 2008), metallurgy (Shi *et al*., 1999), and insurance risk (Brazouskas and Kleefeld, 2009). In this case some care must be taken when considering the analytic bias correction.

Specifically, recall the requirement noted at the end of section 3, that $\xi > -1/3$. As $\hat{\xi}$ approaches this threshold the value of the bias-corrected estimator, $\tilde{\xi}$, becomes unbounded. This has implications for both the design of our Monte Carlo experiment and for practical applications. With regard to our experiment, some preliminary simulation evidence suggested that, to be conservative, $\tilde{\xi}$ should be computed only when $\hat{\xi} > -0.2$. Imposing this condition leads to the following "composite" estimators,

$$\begin{aligned} \hat{\tilde{\xi}} &= \tilde{\xi} \quad ; \quad \hat{\xi} > -0.2 \\ &= \hat{\xi} \quad ; \quad \hat{\xi} \le -0.2 \end{aligned}$$

and (54)

$$\begin{aligned} \hat{\tilde{\sigma}} &= \tilde{\sigma} \quad ; \quad \hat{\xi} > -0.2 \\ &= \hat{\sigma} \quad ; \quad \hat{\xi} \le -0.2 \quad . \end{aligned}$$

As alternative ways of dealing with the biases of $\hat{\xi}$ and $\hat{\sigma}$ we have also considered the (parametric) bootstrap-bias-corrected estimator, as well as Zhang's (2007) likelihood moment estimator, and the quasi-Bayesian estimator of Zhang and Stephens (2009). The bootstrap-bias-corrected estimator is obtained as $\breve{\phi} = 2\hat{\phi} - (1/N_B)[\sum_{j=1}^{N_B} \hat{\phi}_{(j)}]$, where $\hat{\phi}_{(j)}$ is the MLE of ϕ obtained from the j^{th} of the N_B (= 1,000) bootstrap samples, and ϕ is either ξ or σ. See Efron (1982, p.33). This estimator is also unbiased to $O(n^{-2})$, but it is known that this reduction in bias often comes at the expense of increased variance. Zhang's likelihood moment estimator, denoted ϕ^{LME}, and the Zhang and Stephens estimator, denoted ϕ^{ZS}, are computed using the R code supplied by those authors. These two estimators are included in this study given their favourable performances (in terms of both bias and efficiency) in the simulation study reported by Zhang and Stephens (2009). In particular, they dominated the MLE and the MOM and PWM estimators proposed by Hosking and Wallis (1987), so the latter two estimators are not considered here.

Table 1(a): Percentage biases [and MSEs] – shape parameter

n	$\%Bias(\hat{\xi})$ [$\%MSE(\hat{\xi})$]	$\%Bias(\hat{\tilde{\xi}})$ [$\%MSE(\hat{\tilde{\xi}})$]	$\%Bias(\breve{\xi})$ [$\%MSE(\breve{\xi})$]	$\%Bias(\xi^{LME})$ [$\%MSE(\xi^{LME})$]	$\%Bias(\xi^{ZS})$ [$\%MSE(\xi^{ZS})$]
			ξ = 0.4		
50	-11.798 [30.327]	1.016 [22.369]	0.806 [27.635]	-10.270 [27.135]	0.069 [24.686]
100	-5.865 [13.526]	-0.016 [11.694]	0.003 [12.915]	-5.264 [12.802]	0.327 [12.142]
200	-3.025 [6.452]	-0.227 [6.028]	-0.202 [6.306]	-2.772 [6.290]	0.029 [6.094]
500	-1.107 [2.491]	-0.011 [2.428]	-0.009 [2.472]	-1.015 [2.469]	0.093 [2.435]
			ξ = 0.2		
50	-26.267 [98.886]	3.386 [71.104]	1.746 [85.687]	-19.184 [85.649]	5.279 [78.349]
100	-12.531 [42.339]	1.328 [33.289]	0.532 [39.410]	-9.173 [39.553]	3.988 [37.506]
200	-5.969 [19.456]	0.420 [17.397]	0.316 [18.775]	-5.021 [19.352]	1.732 [18.516]
500	-2.474 [7.451]	-0.016 [7.141]	-0.019 [7.344]	-1.844 [7.539]	0.872 [7.249]
			ξ = 0.1		
50	-56.502 [358.213]	3.986 [283.990]	3.316 [299.044]	-37.213 [304.59]	14.753 [278.936]
100	-27.568 [150.287]	3.784 [112.202]	0.511 [136.723]	-18.111 [140.050]	9.909 [131.547]
200	-12.994 [68.113]	1.456 [58.050]	0.598 [64.919]	-8.435 [67.579]	6.075 [63.629]
500	-5.042 [25.247]	0.393 [23.816]	0.292 [24.762]	-3.148 [25.932]	2.711 [24.515]
			ξ = -0.1		
50	-64.681 [309.862]	-27.856 [334.545]	5.419 [233.657]	-33.501 [253.223]	22.755 [231.229]
100	-32.123 [121.356]	-4.402 [122.503]	1.086 [104.510]	-15.788 [113.853]	14.044 [103.929]
200	-16.442 [52.232]	1.145 [48.099]	0.191 [47.749]	-7.589 [53.413]	7.647 [47.682]
500	-7.011 [18.573]	0.948 [16.416]	-0.195 [17.747]	-3.017 [20.341]	3.147 [17.724]

Table 1(a) (continued): Percentage biases [and MSEs] – shape parameter

n	$\%Bias(\hat{\xi})$ [$\%MSE(\hat{\xi})$]	$\%Bias(\hat{\hat{\xi}})$ [$\%MSE(\hat{\hat{\xi}})$]	$\%Bias(\breve{\xi})$ [$\%MSE(\breve{\xi})$]	$\%Bias(\xi^{LME})$ [$\%MSE(\xi^{LME})$]	$\%Bias(\xi^{ZS})$ [$\%MSE(\xi^{ZS})$]
			ξ = -0.15		
50	-45.475 [135.604]	-29.834 [148.169]	3.575 [98.001]	-22.149 [109.576]	15.594 [99.272]
100	-22.531 [51.776]	-9.709 [56.667]	0.678 [43.771]	-10.217 [48.365]	9.609 [43.811]
200	-11.553 [21.836]	-2.507 [23.210]	0.181 [19.705]	-4.769 [22.578]	5.334 [19.798]
500	-4.671 [7.573]	0.329 [7.539]	0.213 [7.202]	-1.558 [8.569]	2.523 [7.229]
			ξ = -0.2		
50	-35.708 [76.073]	-30.442 [80.522]	3.090 [52.506]	-16.095 [60.185]	12.257 [54.369]
100	-18.025 [28.039]	-13.347 [30.718]	0.245 [23.153]	-7.413 [26.233]	7.164 [23.377]
200	-9.146 [11.625]	-5.887 [12.959]	0.176 [10.342]	-3.344 [12.15]	4.119 [10.460]
500	-3.796 [3.961]	-2.068 [4.418]	0.158 [3.729]	-1.052 [4.588]	1.925 [3.755]

Table 1(b): Percentage biases [and MSEs] – scale parameter

n	$\%Bias(\hat{\sigma})$ [$\%MSE(\hat{\sigma})$]	$\%Bias(\hat{\hat{\sigma}})$ [$\%MSE(\hat{\hat{\sigma}})$]	$\%Bias(\breve{\sigma})$ [$\%MSE(\breve{\sigma})$]	$\%Bias(\sigma^{LME})$ [$\%MSE(\sigma^{LME})$]	$\%Bias(\sigma^{ZS})$ [$\%MSE(\sigma^{ZS})$]
			ξ = 0.4		
50	5.770 [7.316]	-1.863 [4.069]	-0.634 [6.113]	4.950 [6.662]	0.697 [5.787]
100	2.879 [3.149]	-0.203 [2.432]	0.005 [2.895]	2.596 [3.025]	0.323 [2.793]
200	1.398 [1.485]	0.004 [1.323]	0.039 [1.428]	1.288 [1.458]	0.156 [1.399]
500	0.557 [0.572]	0.028 [0.547]	0.034 [0.563]	0.519 [0.568]	0.074 [0.558]
			ξ = 0.2		
50	5.993 [6.484]	-2.401 [3.559]	-0.679 [5.234]	4.374 [5.708]	-0.561 [4.916]
100	2.794 [2.731]	-0.679 [1.887]	-0.165 [2.479]	2.092 [2.585]	-0.562 [2.372]
200	1.299 [1.286]	-0.184 [1.097]	-0.092 [1.230]	1.118 [1.282]	-0.248 [1.215]
500	0.537 [0.496]	-0.010 [0.468]	0.000 [0.488]	0.416 [0.499]	-0.135 [0.482]

Table 1(b) (continued): Percentage biases [and MSEs] – scale parameter

n	$\%Bias(\hat{\sigma})$ [$\%MSE(\hat{\sigma})$]	$\%Bias(\hat{\tilde{\sigma}})$ [$\%MSE(\hat{\tilde{\sigma}})$]	$\%Bias(\breve{\sigma})$ [$\%MSE(\breve{\sigma})$]	$\%Bias(\sigma^{LME})$ [$\%MSE(\sigma^{LME})$]	$\%Bias(\sigma^{ZS})$ [$\%MSE(\sigma^{ZS})$]
			***ξ* = 0.1**		
50	6.147 [6.076]	-2.054 [3.790]	-0.782 [4.780]	4.027 [5.234]	-1.167 [4.501]
100	2.948 [2.616]	-0.919 [1.704]	-0.129 [2.346]	1.986 [2.460]	-0.835 [2.245]
200	1.383 [1.203]	-0.256 [0.966]	-0.067 [1.142]	0.942 [1.188]	-0.531 [1.114]
500	0.566 [0.455]	-0.018 [0.423]	0.006 [0.446]	0.387 [0.462]	-0.211 [0.440]
			***ξ* = -0.1**		
50	6.825 [5.577]	2.756 [5.052]	-1.031 [4.012]	3.554 [4.566]	-1.974 [3.916]
100	3.304 [2.292]	0.327 [1.952]	-0.149 [1.984]	1.706 [2.132]	-1.294 [1.918]
200	1.570 [1.032]	-0.289 [0.861]	-0.106 [0.959]	0.733 [1.025]	-0.828 [0.942]
500	0.693 [0.386]	-0.129 [0.336]	0.019 [0.374]	0.322 [0.401]	-0.321 [0.370]
			***ξ* = -0.15**		
50	6.932 [5.441]	4.689 [5.126]	-1.276 [3.797]	3.306 [4.403]	-2.229 [3.779]
100	3.304 [2.211]	1.475 [2.069]	-0.276 [1.901]	1.507 [2.046]	-1.477 [1.846]
200	1.642 [0.994]	0.330 [0.934]	-0.114 [0.918]	0.684 [0.989]	-0.874 [0.904]
500	0.675 [0.372]	-0.064 [0.348]	-0.042 [0.359]	0.244 [0.390]	-0.400 [0.356]
			***ξ* = -0.2**		
50	7.214 [5.432]	6.590 [5.171]	-1.450 [3.619]	3.163 [4.274]	-2.350 [3.678]
100	3.530 [2.145]	2.872 [2.090]	-0.200 [1.815]	1.476 [1.972]	-1.451 [1.772]
200	1.722 [0.958]	1.219 [0.960]	-0.123 [0.878]	0.634 [0.956]	-0.906 [0.868]
500	0.726 [0.356]	0.436 [0.365]	-0.042 [0.342]	0.221 [0.377]	-0.412 [0.340]

Table 1 summarizes the performances of the various estimators in terms of *percentage* biases and MSE's, the former being defined as $100\times$(Bias / | ξ |) and $100\times$(Bias / | σ |), and the latter as $100\times$(MSE / ξ^2) and $100\times$(MSE / σ^2). We see that the original MLEs of the shape and scale parameters are negatively and positively biased, respectively, regardless of the sign of ξ. The percentage bias of $\hat{\xi}$ decreases in absolute value as the true absolute value of the shape parameter increases. This absolute bias is slightly larger for negative values of ξ than for positive values of this parameter in corresponding situations. On the other hand the percentage bias of $\hat{\sigma}$ is relatively robust to changes in the magnitude and sign of the shape parameter. Of course, these absolute biases decline monotonically as the sample size increases. All of these observations are consistent with the results in Table 2 of Hosking and Wallis (1987, p.343), who report that they had difficulties with their Newton-Raphson maximization algorithm for small sample sizes. Moreover, the numerical values of our biases for $\hat{\xi}$ and $\hat{\sigma}$ are very close to those of Hosking and Wallis, once account is taken of the fact that our ξ corresponds to their $-k$, and that they report actual (rather than percentage) biases.

The analytic bias corrections, in the form of $\tilde{\hat{\xi}}$ and $\tilde{\hat{\sigma}}$, perform extremely well in all cases, and generally reduce the percentage biases by at least an order of magnitude when $\xi > 0$. For $\xi < 0$ these bias reductions are still substantial, though less so as ξ becomes increasingly negative. In some cases there is an "over-correction" when the bias correction is applied, with the percentage bias changing sign. This can been seen in Table 1 when $\xi = 0.1$, for all values of n considered. Similar results are reported by Cribari-Neto and Vaconcellos (2002) in the case of the beta distribution, Giles (2011) for the half-logistic distribution, and Giles *et al*. (2011) for the Lomax distribution.

It is extremely encouraging that the reduction in the relative biases for the (bias-corrected) estimators of both ξ and σ is accompanied by a small reduction in relative mean squared error when $\xi > 0$. With only two exceptions (for large n when $\xi = -0.2$), the same is true for the estimators of σ when $\xi < 0$, and in the exceptional cases the %MSE is essentially unchanged by the bias correction. Analytically bias-adjusting the MLE for ξ, when that parameter is negative, can affect the %MSE either favourably or unfavourably, but only very modestly.

The results of bootstrap bias-correcting $\hat{\xi}$ and $\hat{\sigma}$ are also very satisfactory, as absolute percentage biases are reduced, and so are %MSE's, in all of the cases considered in Table 1. The same is true for

Zhang's likelihood moment estimator, and the estimator proposed by Zhang and Stephens. Consistent with the latter authors' results, ϕ^{ZS} dominates ϕ^{LME} with respect to both bias and efficiency when the scale parameter is positive, and in many of the cases considered when it is negative. The bootstrap bias-corrected estimator typically dominates Zhang and Stephens' estimator in terms of percentage bias – often by an order of magnitude. For positive values of the shape parameter, the %MSE of the bootstrap estimator is very close to that of the likelihood moment estimator; while it is very close to that of Zhang and Stephens' estimator when $\xi < 0$. Overall, the bootstrap estimator is dominant among these three. However, when both bias and efficiency are taken into account, together with computational cost, our composite bias-corrected estimators perform very credibly, especially when the shape parameter is positive.

Making comparisons between the effectiveness of our analytical bias correction and that of the other three estimators is complicated by the fact that we are proposing "composite" estimators, $\hat{\tilde{\xi}}$ and $\hat{\tilde{\sigma}}$, given in (54), in Table 1. If $\hat{\xi} > -0.2$, then the appropriate comparison is that between the "pure" analytically bias-adjusted estimators ($\tilde{\xi}$ and $\tilde{\sigma}$), and the alternative estimators (bootstrap-corrected, likelihood moment, and Zhang-Stephens). This is facilitated in Table 2 with the measures $\Delta_{PB}(\breve{\xi},\tilde{\xi})$ and $\Delta_{PM}(\breve{\xi},\tilde{\xi})$, where $\Delta_{PB}(\breve{\xi},\tilde{\xi})$ = $|\%Bias(\breve{\xi})|$-$|\%Bias(\tilde{\xi})|$, and $\Delta_{PM}(\breve{\xi},\tilde{\xi})$ = $|\%MSE(\breve{\xi})|$-$|\%MSE(\tilde{\xi})|$. Corresponding measures are defined for a comparison between $\breve{\sigma}$ and $\tilde{\sigma}$, and for comparisons between our composite estimator and each of the estimators proposed by Zhang (2007) and Zhang and Stephens (2009). To provide an alternative perspective to that given in Table 1, the values in Table 2 have been computed only for those Monte Carlo replications in which $\hat{\xi} > -0.2$. For example, when $\xi = -0.2$, and n = 100, only 37% of the 50,000 Monte Carlo replications resulted in an analytic correction. The corresponding percentages when $\xi = -0.1$ and n = 100, and when $\xi = 0.2$ and n = 100, are 75.3% and 99.7% respectively.

In Table 2, positive values imply that the second-named estimator dominates the first-named estimator, in terms of either bias or MSE. We see that in terms of the latter criterion, our pure Cox-Snell correction always dominates the other three bias corrections in terms of %MSE. Comparing the various bias corrected estimators of both of the parameters, in terms of percentage bias reduction, we see that although the results are somewhat "mixed", the dominant pattern is that the order of preference is $\breve{\phi} \succ \tilde{\phi} \succ \phi^{ZS} \succ \phi^{LME}$ (for $\phi = \xi, \sigma$) when $\xi > 0$, and $\phi^{LME} \succ \breve{\phi} \succ \tilde{\phi} \succ \phi^{ZS}$ when $\xi < 0$.

When the MSE reduction and the computational simplicity are taken into account, the Cox-Snell analytic bias correction can be recommended unless the MLE for the shape parameter is “too close” to its lower threshold value of -1/3. The latter constraint ensures the validity of the Cox-Snell bias correction, and it is more than sufficient for the likelihood function to be regular. So, the analytic correction is unlikely to be applied when the regularity conditions are violated if our recommendation is followed. In contrast, the bootstrap bias correction can be applied even when the regularity conditions fail, resulting in modified MLEs that lack their usual desirable asymptotic properties.

Finally, the first and fifth columns of results in Table 2 should not affect the justification for our earlier recommendation to use the “composite” estimators, $\hat{\tilde{\xi}}$ and $\hat{\tilde{\sigma}}$. With two (small *n*) exceptions the values of $\Delta_{PB}(\hat{\xi},\tilde{\xi})$ and $\Delta_{PB}(\hat{\sigma},\tilde{\sigma})$ are positive when $\xi > 0$. However, they are negative (with one exception) when the shape parameter is negative. The latter result is not surprising, and is due to the decision rule of bias-correcting only when $\hat{\xi} > -0.2$. Since the bias of $\hat{\xi}$ is negative, bias-correction tends to make the estimate larger, and when $\xi < 0$ it is likely that only estimates that are already too large will meet the decision criteria.

Table 2: Comparisons between percentage biases [and MSEs]

n	$\Delta_{PB}(\hat{\xi},\tilde{\xi})$ $\Delta_{PM}(\hat{\xi},\tilde{\xi})$	$\Delta_{PB}(\xi^{ZS},\tilde{\xi})$ $\Delta_{PM}(\xi^{ZS},\tilde{\xi})$	$\Delta_{PB}(\xi^{LME},\tilde{\xi})$ $\Delta_{PM}(\xi^{LME},\tilde{\xi})$	$\Delta_{PB}(\breve{\xi},\tilde{\xi})$ $\Delta_{PM}(\breve{\xi},\tilde{\xi})$	$\Delta_{PB}(\hat{\sigma},\tilde{\sigma})$ $\Delta_{PM}(\hat{\sigma},\tilde{\sigma})$	$\Delta_{PB}(\sigma^{ZS},\tilde{\sigma})$ $\Delta_{PM}(\sigma^{ZS},\tilde{\sigma})$	$\Delta_{PB}(\sigma^{LME},\tilde{\sigma})$ $\Delta_{PM}(\sigma^{LME},\tilde{\sigma})$	$\Delta_{PB}(\breve{\sigma},\tilde{\sigma})$ $\Delta_{PM}(\breve{\sigma},\tilde{\sigma})$
				ξ = 0.4				
50	8.519 [8.012]	-1.151 [2.953]	7.136 [5.287]	-0.342 [5.727]	3.031 [3.269]	-1.957 [1.881]	2.301 [2.736]	-1.324 [2.174]
100	5.850 [1.831]	0.340 [0.452]	5.250 [1.112]	0.017 [1.224]	2.663 [0.717]	0.108 [0.362]	2.381 [0.595]	-0.208 [0.464]
200	2.798 [0.424]	-0.198 [0.066]	2.545 [0.262]	-0.025 [0.278]	1.394 [0.161]	0.152 [0.075]	1.284 [0.135]	0.035 [0.105]
500	1.096 [0.062]	0.082 [0.007]	1.004 [0.041]	-0.002 [0.044]	0.530 [0.024]	0.046 [0.011]	0.492 [0.021]	0.006 [0.016]
				ξ = 0.2				
50	3.524 [28.928]	-0.234 [17.387]	-2.437 [21.612]	-3.299 [22.554]	-0.214 [3.045]	-2.513 [1.965]	-1.508 [2.623]	-2.268 [2.196]
100	9.957 [9.076]	2.586 [4.568]	6.664 [6.587]	-0.851 [6.363]	1.888 [0.846]	-0.139 [0.503]	1.204 [0.715]	-0.530 [0.606]
200	6.429 [2.182]	1.567 [0.883]	4.839 [1.719]	-0.105 [1.379]	1.372 [0.197]	0.187 [0.105]	1.055 [0.173]	-0.092 [0.133]
500	2.458 [0.310]	0.856 [0.108]	1.828 [0.397]	0.003 [0.203]	0.526 [0.028]	0.124 [0.014]	0.406 [0.031]	-0.010 [0.020]

Table 2 (continued): Comparisons between percentage biases [and MSEs]

n	$\Delta_{PB}(\hat{\xi},\widetilde{\xi})$ $\Delta_{PM}(\hat{\xi},\widetilde{\xi})$	$\Delta_{PB}(\xi^{ZS},\widetilde{\xi})$ $\Delta_{PM}(\xi^{ZS},\widetilde{\xi})$	$\Delta_{PB}(\xi^{LME},\widetilde{\xi})$ $\Delta_{PM}(\xi^{LME},\widetilde{\xi})$	$\Delta_{PB}(\breve{\xi},\widetilde{\xi})$ $\Delta_{PM}(\breve{\xi},\widetilde{\xi})$	$\Delta_{PB}(\hat{\sigma},\widetilde{\sigma})$ $\Delta_{PM}(\hat{\sigma},\widetilde{\sigma})$	$\Delta_{PB}(\sigma^{ZS},\widetilde{\sigma})$ $\Delta_{PM}(\sigma^{ZS},\widetilde{\sigma})$	$\Delta_{PB}(\sigma^{LME},\widetilde{\sigma})$ $\Delta_{PM}(\sigma^{LME},\widetilde{\sigma})$	$\Delta_{PB}(\breve{\sigma},\widetilde{\sigma})$ $\Delta_{PM}(\breve{\sigma},\widetilde{\sigma})$
					ξ = 0.1			
50	-22.705 [81.739]	1.279 [69.105]	-37.820 [67.612]	-8.608 [74.607]	-3.638 [2.518]	-2.335 [1.824]	-5.170 [2.248]	-2.516 [1.975]
100	13.637 [38.653]	5.398 [24.985]	4.678 [32.372]	-3.871 [28.580]	1.020 [0.926]	-0.190 [0.619]	0.125 [0.821]	-0.874 [0.701]
200	11.269 [10.067]	4.608 [5.658]	6.726 [9.643]	-0.866 [6.922]	1.097 [0.226]	0.273 [0.139]	0.658 [0.213]	-0.190 [0.167]
500	4.649 [1.431]	2.317 [0.699]	2.755 [2.115]	-0.102 [0.945]	0.548 [0.033]	0.193 [0.018]	0.370 [0.039]	-0.012 [0.023]
					ξ = -0.1			
50	-60.477 [-40.537]	17.752 [72.851]	-40.174 [-3.083]	2.987 [40.594]	-6.682 [0.861]	0.324 [1.573]	-4.917 [1.117]	-0.323 [1.477]
100	-36.826 [-1.524]	7.220 [25.952]	-24.198 [14.941]	-5.049 [14.135]	-3.954 [0.451]	0.216 [0.613]	-2.834 [0.581]	-0.022 [0.532]
200	-19.966 [4.692]	3.672 [8.827]	-12.334 [11.528]	-3.660 [6.289]	-2.111 [0.195]	0.199 [0.207]	-1.421 [0.252]	-0.260 [0.161]
500	1.958 [2.196]	2.031 [1.795]	-1.879 [4.462]	-1.307 [1.684]	0.127 [0.052]	0.172 [0.041]	-0.225 [0.072]	-0.168 [0.036]

Table 2 (continued): Comparisons between percentage biases [and MSEs]

n	$\Delta_{PB}(\hat{\xi},\tilde{\xi})$ $\Delta_{PM}(\hat{\xi},\tilde{\xi})$	$\Delta_{PB}(\xi^{ZS},\tilde{\xi})$ $\Delta_{PM}(\xi^{ZS},\tilde{\xi})$	$\Delta_{PB}(\xi^{LME},\tilde{\xi})$ $\Delta_{PM}(\xi^{LME},\tilde{\xi})$	$\Delta_{PB}(\breve{\xi},\tilde{\xi})$ $\Delta_{PM}(\breve{\xi},\tilde{\xi})$	$\Delta_{PB}(\hat{\sigma},\tilde{\sigma})$ $\Delta_{PM}(\hat{\sigma},\tilde{\sigma})$	$\Delta_{PB}(\sigma^{ZS},\tilde{\sigma})$ $\Delta_{PM}(\sigma^{ZS},\tilde{\sigma})$	$\Delta_{PB}(\sigma^{LME},\tilde{\sigma})$ $\Delta_{PM}(\sigma^{LME},\tilde{\sigma})$	$\Delta_{PB}(\breve{\sigma},\tilde{\sigma})$ $\Delta_{PM}(\breve{\sigma},\tilde{\sigma})$
					ξ = -0.15			
50	-32.978 [-26.492]	20.615 [45.917]	-17.855 [-1.117]	10.080 [25.725]	-4.728 [0.665]	2.196 [1.885]	-2.833 [1.084]	1.477 [1.711]
100	-22.074 [-8.420]	7.988 [13.769]	-12.989 [3.071]	-0.358 [5.551]	-3.149 [0.244]	0.985 [0.665]	-1.979 [0.450]	-0.022 [0.532]
200	-12.999 [-1.974]	3.295 [4.059]	-7.452 [3.198]	-1.737 [1.581]	-1.886 [0.086]	0.446 [0.209]	-1.156 [0.187]	-0.260 [0.161]
500	-5.839 [0.040]	1.260 [0.815]	-3.144 [1.871]	-1.040 [0.447]	-0.864 [0.028]	0.183 [0.043]	-0.505 [0.067]	-0.168 [0.036]
					ξ = -0.2			
50	-15.583 [-13.166]	25.801 [41.846]	-2.316 [7.532]	17.249 [26.717]	-1.846 [0.775]	4.952 [2.579]	0.293 [1.402]	4.153 [2.317]
100	-12.615 [-7.224]	10.584 [12.578]	-4.136 [3.103]	4.018 [5.689]	-1.774 [0.146]	2.305 [0.892]	-0.359 [0.485]	1.278 [0.673]
200	-8.091 [-3.313]	4.428 [3.639]	-3.066 [1.467]	0.573 [1.135]	-1.248 [-0.004]	1.052 [0.287]	-0.388 [0.163]	0.351 [0.187]
500	-3.980 [-1.054]	1.504 [0.718]	-1.420 [0.754]	-0.246 [0.091]	-0.668 [-0.021]	0.377 [0.060]	-0.218 [0.047]	0.032 [0.031]

5. Empirical examples

We present some empirical examples based on a variety of real data-sets to illustrate the practical impact of using the analytical bias adjustment for the MLEs of the parameters of the GPD. The first two examples simply relate to estimates that are already reported in the literature, and are chosen for their relatively small sample sizes and the fact that they involve both positive and negative estimates of the shape parameter.

5.1 Rainfall data

Van Montfort and Witter (1986) use the GPD to model Dutch rainfall data. The cases that they consider involve a range of sample sizes, and in Table 3 we report their MLEs for their three smallest-sized samples. The corresponding analytically bias-adjusted estimates are also reported, with bootstrapped standard errors based on 20,000 bootstrap samples. We see that the estimates of the shape parameter are quite sensitive to the bias adjustment, with a change in sign occurring when n = 87.

The bootstrapped standard errors for $\tilde{\xi}$ and $\tilde{\sigma}$ describe the precision of these estimates. We can also test if the bias-adjusted estimates are *significantly* different from the original MLEs, taking account of the fact that the sampling distributions of $\tilde{\xi}$ and $\tilde{\sigma}$ are highly non-normal, especially at these sample sizes. The bootstrap p-values (p) in Table 3 facilitate this. Consider the results for n = 83, for example. There is a probability of only 0.6% that a value as "extreme" as the estimate based on $\hat{\xi}$ will occur, conditional on the sampling distribution of $\tilde{\xi}$. In this sense, one concludes that the bias-adjusted estimate of ξ is significantly different from the unadjusted estimate. With a nominal significance level of 2%, say, the same result holds for all of the other parameter estimates in Table 3.

Table 3: Dutch rainfall data [1]

n	$\hat{\xi}$ (a.s.e.)	$\tilde{\xi}$ (b.s.e.)	p	$\hat{\sigma}$ (a.s.e.)	$\tilde{\sigma}$ (b.s.e.)	p
83	0.091	0.225	0.006	9.142	7.650	0.016
	(0.123)	(0.117)		(1.508)	(1.027)	
84	0.002	0.160	0.011	7.074	5.673	0.010

	(0.129)	(0.108)		(1.200)	(0.746)	
87	-0.031	0.127	0.016	4.994	3.990	0.012
	(0.116)	(0.105)		(0.805)	(0.518)	

[1.] "a.s.e." denotes asymptotic standard error. "b.s.e." denotes bootstrapped standard error, based on 20,000 bootstrap samples.

5.2 Experimental steels data

Shi *et al*. (1999) use maximum likelihood estimation to fit the GPD to data for "inclusions" collected on the surfaces of cold crucible re-melted steels. In Table 4 we reproduce the results from their Table 3 for a selection of thresholds, *u*, above which the GPD was fitted by MLE using the associated exceedances. In all cases, $\hat{\xi} > -0.2$, consistent with the rule of thumb suggested in conjunction with equation (54) above. The number of exceedances (*n*) were reported only graphically by Shi *et al.* (1999) in their Figure 7, and our "reverse engineered" values are provided in Table 4 below.

Using these data we have bias-adjusted the MLEs for the parameters of the GPD, for two steel types – A and B. We see in Table 4 that correcting for bias can alter the point estimates dramatically. This is especially so in the case of the shape parameter, where there are many instances of sign changes as we move from $\hat{\xi}$ to $\tilde{\xi}$. The effect of these changes on the estimated survival function for the GPD is illustrated in Figure 1, for the case of u = 2.8 for Steel B. We see, for example, that the 99th percentile of the distribution increases from 4μm to 5μm as a result of bias-correcting the MLEs. The numerical effect on the survival function is substantial.

Table 4: Experimental steels data[1]

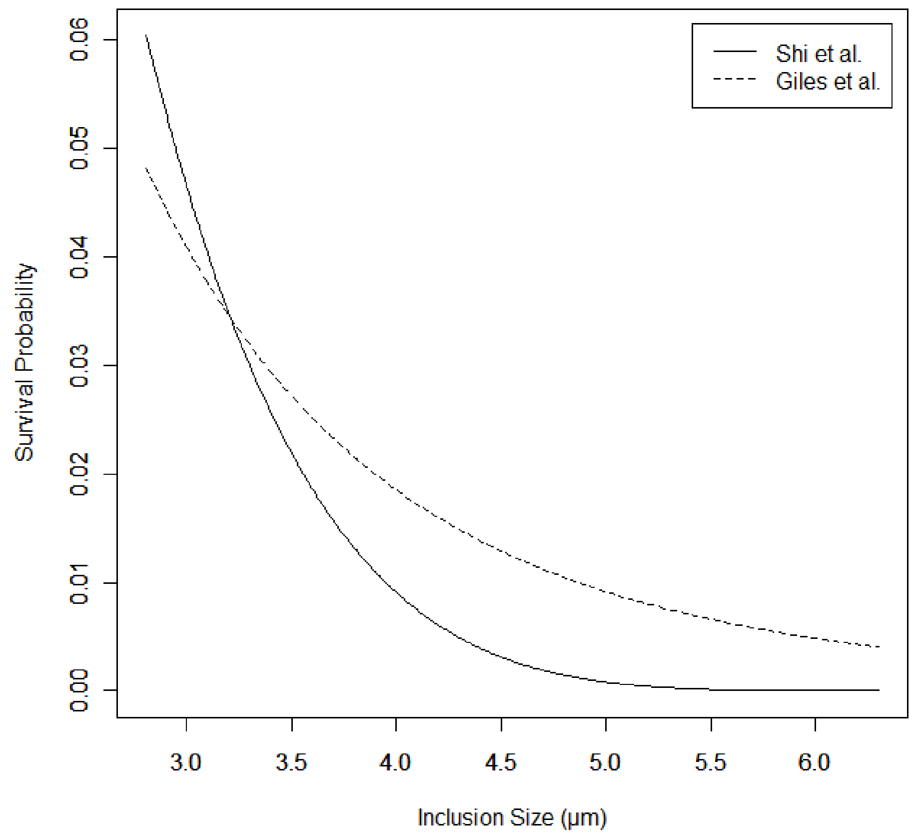
Steel A					Steel B				
u (μm) [n]	$\hat{\xi}$	$\hat{\sigma}$	$\tilde{\xi}$	$\tilde{\sigma}$	u (μm) [n]	$\hat{\xi}$	$\hat{\sigma}$	$\tilde{\xi}$	$\tilde{\sigma}$

2.0	-0.098	1.616	0.034	1.350	2.0	-0.195	1.420	0.007	1.049
[78]	(0.123)	(0.275)	(0.125)	(0.203)	[72]	(0.124)	(0.250)	(0.130)	(0.167)
2.2	-0.055	1.455	0.057	1.258	2.2	-0.097	1.150	-0.011	1.038
[70]	(0.135)	(0.269)	(0.218)	(0.422)	[69]	(0.133)	(0.212)	(0.230)	(0.303)
2.4	-0.055	1.450	0.071	1.230	2.4	-0.113	1.160	0.004	1.005
[62]	(0.146)	(0.289)	(0.226)	(0.418)	[55]	(0.152)	(0.246)	(0.245)	(0.305)
2.6	-0.062	1.460	0.099	1.175	2.6	-0.060	1.050	0.025	0.957
[50]	(0.168)	(0.334)	(0.242)	(0.414)	[48]	(0.172)	(0.248)	(0.245)	(0.280)
2.8	-0.036	1.370	0.116	1.123	2.8	-0.195	1.300	0.153	0.728
[45]	(0.180)	(0.339)	(0.251)	(0.398)	[36]	(0.213)	(0.382)	(0.244)	(0.173)
3.0	-0.060	1.440	0.136	1.098	3.0	-0.175	1.230	0.099	0.819
[40]	(0.196)	(0.386)	(0.262)	(0.340)	[31]	(0.250)	(0.441)	(0.283)	(0.263)

[1.] Bootstrapped standard errors, based on 20,000 bootstrap samples, are reported in parentheses. Standard errors are not reported for the original MLEs by Shi *et al.* (1999).

Figure 1: Survival functions of GPD for experimental Steel data

(Steel B; *u* = 2.8)



5.3 Stock market data

Our first new application uses the daily returns (log-differences) of the closing values for the Dow-Jones Industrial Average share price index between 6 July 2008 and 7 July 2009. Positive and negative daily returns were analyzed separately using the peaks-over-threshold technique, to allow for possible asymmetries. Both series are stationary and serially independent. Summary statistics appear in part (a) of Table 5. Using the graphical aids in the POT package (Ribatet, 2007) for *R* we determined a threshold of u = 3.0% (2%) for the positive (negative) returns, resulting in 24 (66) "exceedances" for the positive (negative) returns.

In addition to exploring the numerical impact of bias adjustment on the MLEs of the parameters, we also consider the implications for two risk measures – "value-at-risk" (VaR) and "expected shortfall" (ES) – that are computed using the estimated parameters. Conceptually, VaR_p is the $(1\text{-}p)^{th}$ quantile of the distribution. If $VaR_{0.01}$ = 5, say, there is a 1% probability that the data will exceed the value of 5. Similarly, $ES_{0.01}$ is the (conditional) mean of the data values that exceed the $VaR_{0.01}$. When the GPD is used to model the "exceedances" (defined as $y_i = x_i - u$, where x_i denotes an original observation) above the selected threshold, u, in the peaks-over-threshold method (*e.g.*, Coles, 2001, chap. 4) it is readily established that

$$VaR_p = u + \frac{\sigma}{\xi}\left(\left(\frac{N}{n}p\right)^{-\xi} - 1\right)$$

and

$$ES_p = (VaR_p + \sigma - \xi u)/(1-\xi) \, ,$$

where p is the desired tail probability, N is the original sample size, and n is the number of exceedances (McNeill, 1997; Bi and Giles, 2009).

The MLE results and associated estimated risk measures are given in part (b) of Table 5. We see that the shape and scale parameter estimates are under-stated and over-stated respectively prior to bias adjustment. The implication for the risk measures of failing to bias-adjust the parameter estimates is that they are all too conservative. This is especially so in the case of the estimated expected shortfall for positive returns.

By way of comparison, Coles (2001, pp.86-90) analyzes a sample of 1,303 daily returns for the Dow Jones index. He does not separate positive returns from negative returns, and determines that a threshold value of u = 2% is appropriate, yielding n = 37 exceedances. His MLEs (with standard errors) are $\hat{\xi}$ = 0.288 (0.258) and $\hat{\sigma}$ = 0.495 (0.150). Using these values we can obtain the bias-adjusted parameter estimates, $\tilde{\xi}$ = 0.163 (0.218) and $\tilde{\sigma}$ = 0.570 (0.156). Although the numerical impact of bias adjustment on the parameter estimates is quite marked, this does not carry through to the estimates of VaR and ES. Based on Coles' original MLEs, these estimates are 2.60% and 3.54% respectively, and they change to 2.65% and 3.46% when the MLEs are adjusted for bias.

5.4 Billion dollar weather disasters

Our final example involves fitting a GPD to all of the 58 weather-related disasters in the U.S.A. between 1980 and 2003 that resulted in damages in excess of $1Billion. The data are from Ross and Lott (2003), and are in real 2002 billions of dollars. The summary statistics for the data appear in Table 6, together with the MLEs for the GPD parameters and the estimated 5% values-at-risk and expected shortfalls.

Table 5: Dow-Jones data (6 July 2008 – 7 July 2009)

(a) Summary statistics

	Positive returns (%)		Negative returns (%)	
	Full sample	Exceedances	Full sample	Exceedances
	$N = 247$	$n = 24$	$N = 258$	$n = 66$
Mean	1.331	3.192	-1.471	-3.187
Median	0.803	2.776	-1.113	-2.673
Maximum (Minimum)	10.508	10.508	(-8.201)	(-8.201)
Standard deviation	1.462	1.518	1.460	1.387
Coefficient of variation (%)	109.8971	47.550	99.239	43.514
Skewness	2.858	2.750	-1.911	-1.873
Kurtosis	15.19821	12.296	7.670	6.323

(b) Estimation results[1]

	Positive returns	Negative returns
$\hat{\xi}$ (a.s.e.)	0.388 (0.335)	0.100 (0.167)
$\hat{\sigma}$ (a.s.e.)	1.080 (0.415)	1.272 (0.263)
$\tilde{\xi}$ (b.s.e.)	0.497 (0.343)	0.170 (0.141)
$\tilde{\sigma}$ (b.s.e.)	0.942 (0.406)	1.171 (0.194)
$V\hat{a}R_{0.01}$ ($V\tilde{a}R_{0.01}$)	6.94% (6.97%)	6.87% (7.06%)
$E\hat{S}_{0.01}$ ($E\tilde{S}_{0.01}$)	11.20% (12.77%)	8.82% (9.51%)

1. "a.s.e." denotes asymptotic standard error. "b.s.e." denotes bootstrapped standard error, based on 20,000 bootstrap samples.

Table 6: Weather disasters (1980 - 2003)

(a) Summary statistics (billions of 2002 $'s)

n	58	Order statistics:	$1^{st} - 4^{th}$	1.1
Mean	6.03		54^{th}	13.9
Median	2.45		55^{th}	26.7
Standard deviation	11.02		56^{th}	35.6
Skewness	3.70		57^{th}	48.4
Kurtosis	16.59		58^{th}	61.6

(b) Estimation results[1]

$\hat{\xi}$ (a.s.e.)	0.736	(0.223)	$\tilde{\xi}$ (b.s.e.)	0.803	(0.220)
$\hat{\sigma}$ (a.s.e.)	1.709	(0.410)	$\tilde{\sigma}$ (b.s.e.)	1.569	(0.352)
$V\hat{a}R_{0.05}$	$19.7 Billion		$V\tilde{a}R_{0.05}$	$20.7 Billion	
$E\hat{S}_{0.05}$	$78.3 Billion		$E\tilde{S}_{0.05}$	$109.0 Billion	

1. "a.s.e." denotes asymptotic standard error. "b.s.e." denotes bootstrapped standard error, based on 20,000 bootstrap samples.

We see that in this example the estimates of the shape parameter imply that second and higher-order moments of the underlying GPD do not exist. In addition, the effect of bias-correcting the parameter estimates is to increase the 5% value-at-risk by $1 Billion, and the associated expected shortfall by nearly $31 Billion. It should be noted that in this case the threshold for the latter calculations is $1 Billion. So, the interpretation of the bias-corrected VaR is that once the damage bill for a weather-related disaster reaches $1 Billion, there is a 5% probability that it will ultimately reach $20.7 Billion or more. The interpretation of the corresponding ES is that, conditional on the damage bill reaching its VaR value of $20.7 Billion, we can expect that the final bill will reach $109 Billion. The estimated VaR value seems very reasonable when we consider the largest four order statistics given in Table 6 (a). Four (or 6.9%) of the 58 observations exceed $20.7 Billlion.

6. Conclusions

We have derived analytic expressions for the bias to $O(n^{-1})$ of the maximum likelihood estimators of the parameters of the generalized Pareto distribution. These have then been used to bias-correct the original estimators, resulting in modified estimators that are unbiased to order $O(n^{-2})$. Specifically, we have considered "composite" estimators which involve bias correcting in this way only if the MLE of the shape parameter is in a specified range. We find that the negative relative bias of the shape parameter estimator, and the positive relative bias of the scale parameter estimator are each reduced dramatically by using this correction. This reduction is especially noteworthy in the case of the shape parameter. Importantly, these gains are usually obtained with a small reduction in relative mean squared error when the shape parameter is positive, and only very minimal increases in this measure when this parameter is negative, at least for sample sizes of the magnitude likely to be encountered in practice.

Using the bootstrap to bias-correct the maximum likelihood estimators of the parameters is also extremely effective for this distribution. However, on balance it is inferior to the analytic "composite" correction, especially once the effect on mean squared error is considered, and computational costs and robustness are taken into account. Alternative estimators for this distribution's parameters have been proposed by Zhang (2007) and by Zhang and Stephens (2009). Although these are not "bias-adjusted" estimators, they are known to perform well in this respect. However, our simulation results support the use of our analytic bias correction (in its "composite" form) for the MLEs, once bias, efficiency, and computational cost are taken into account.

While reducing the finite-sample bias of the MLEs of the parameters of the GPD is important in its own right, there is also considerable interest in managing the bias of the MLEs of certain functions of these parameters, such as the quantiles of the distribution (see Hosking and Wallis, 1987; Moharram *et al*., 1993, for example). Specifically, in risk analysis we are concerned with value at risk (VaR) and the expected shortfall (ES), both of which are related to these quantiles. These measures are non-linear functions of the shape and scale parameters when the GPD is used in the context of the peaks over threshold method. Work in progress addresses this issue by deriving the Cox-Snell $O(n^{-1})$ biases for the maximum likelihood estimators of VaR and ES themselves, and evaluating the bias-corrected estimators in a manner similar to that adopted in the present paper.

Acknowledgment

We are extremely grateful to Scott Grimshaw for supplying the *R* code that we used to implement his maximum likelihood algorithm, and to Lief Bluck and Kees van Kooten for providing access to the computing resources needed to complete this study in a timely manner. We are also appreciative of the helpful comments, suggestions and questions received from Hugh Chipman, Paul Della-Marta, Ruud Koning, Jacob Schwartz, and participants at the 2009 Hawaii International Conference on Statistics, Mathematics and Related Fields; the 2010 Joint Statistical Meetings; and seminars in the Departments of Economics at the Universities of Auckland and Waikato, and the Department of Mathematics and Statistics at the University of Victoria. The second author acknowledges financial support from King's University College at the University of Western Ontario.

References

Angelini, F. (2002). An analysis of Italian financial data using extreme value theory. *Giornale dell'Istituto Italiano degli Attuari*, LXV, 91-109.

Arnold, B. (1983), *Pareto Distributions*. International Co-operative Publishing House, Fairland, MD.

Bali, T. G. and S. N. Neftci (2003). Disturbing extremal behavior of spot rate dynamics. *Journal of Empirical Finance*, 10, 455-477.

Bartlett, M. S. (1953a). Approximate confidence intervals. *Biometrika*, 40, 12-19.

Bartlett, M. S. (1953b). Approximate confidence intervals II. More than one unknown parameter. *Biometrika*, 40, 306-317.

Bi, G. and D. E. A. Giles (2009). Modelling the financial risk associated with U.S. movie box office earnings. *Mathematics and Computers in Simulation*, 79, 2759-2766.

Brazouskas, V and A. Kleefeld (2009). Robust and efficient fitting of the generalized Pareto distribution with actuarial applications in view. *Insurance: Mathematics and Economics*, 45, 424-435.

Brooks, C., A. D. Clare, J. W. Dalle Molle and G. Persand (2005). A comparison of extreme value theory approaches for determining value at risk. *Journal of Empirical Finance*, 12, 339-352.

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

Castillo, J. and J. Daoudi (2009). Estimation of the generalized Pareto distribution. *Statistics and Probability Letters*, 79, 684-688.

Cebriàn, A. C., M. Denuit and P. Lambert (2003). Generalized Pareto fit to the Society of Actuaries' large claims database. *North American Actuarial Journal*, 18-36.

Chaouche, A. and J-N. Bacro (2006). Statistical inference for the generalized Pareto distribution: Maximum likelihood revisited. *Communications in Statistics - Theory and Methods*, 35, 785-802.

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

Cordeiro, G. M. and R. Klein (1994). Bias correction in ARMA models. *Statistics and Probability Letters*, 19, 169-176.

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

Cramér, H. (1946), *Mathematical Methods of Statistics*. Princeton University Press, Princeton NJ.

Cribari-Neto, F. and K. L. P. Vasconcellos (2002). Nearly unbiased maximum likelihood estimation for the beta distribution. *Journal of Statistical Computation and Simulation*, 72, 107-118.

Davison, A. C. and R. L. Smith (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society, B*, 52, 393-442.

de Zea Bermudez, P. and S. Kotz (2010). Parameter estimation of the generalized Pareto distribution – Part I. *Journal of Statistical Planning and Inference*, 140, 1353-1373.

Dugué, D. (1937). Application des propriétés de la limite au sens du calcul des probabilités à l'étude de diverses questions d'estimation. *Journal de l'École Polytechnique*, 3, 305-372.

Efron, B. (1982), *The Jackknife,the Bootstrap, and Other Resampling Plans*. Society of Industrial Mathematics, Philadalphia PA.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80, 27-38.

Gençay, R and F. Selçuk (2006). Overnight borrowing, interest rates and extreme value theory. *European Economic Review*, 50. 547-563.

Giles, D. E. (2011). Bias reduction for the maximum likelihood estimators of the parameters in the half-logistic distribution. *Communications in Statistics – Theory and Methods*, forthcoming.

Giles, D. E., H. Feng and R. T. Godwin (2011). On the bias of the maximum likelihood estimator for the two-parameter Lomax distribution. *Communications in Statistics – Theory and Methods*, forthcoming.

Gilli, M. and E. Këllizi (2006). An application of extreme value theory for measuring financial risk. *Computational Economics*, 27, 201-228.

Gradshteyn, I. S., Ryzhik, I. W. (1965). *Table of Integrals, Series, and Products* (translation ed. A. Jeffrey), 4th ed. Academic Press, New York.

Grimshaw, S. D. (1993). Computing maximum likelihood estimates for the generalized Pareto distribution. *Technometrics*, 35, 185-191.

Haldane, J. B. S. (1953). The estimation of two parameters from a sample. *Sankhyā*, 12, 313-320.

Haldane, J. B. S. and S. M. Smith (1956). The sampling distribution of a maximum likelihood estimate. *Biometrika*, 43, 96-103.

Holmes, J. D. and W. W. Moriarty (1999). Application of the generalized Pareto distribution to extreme value analysis in wind engineering. *Journal of Wind Engineering and Industrial Aerodynamics*, 83, 1-10.

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

Huyse, L., R. Chen and J. A. Stamatakos (2010). Application of generalized Pareto distribution to constrain uncertainty in peak ground accelerations. *Bulletin of the Seismological Society of America*, 100, 87-101.

Jondeau, E. and M. Rockinger (2003). Testing for differences in the tails of stock-market returns. *Journal of Empirical Finance*, 10, 559-581.

Klüppelberg, C. (2002). Risk management with extreme value theory. Sonderforschungsbereich 386, Paper 270, Institut für Statistick, Ludwig-Maximilians- Universität, München.

Luceño, A. (2006). Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators. *Computational Statistics and Data Analysis*, 51, 904-917.

McNeil, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin*, 27, 117-137.

Moharram, S.H., A. K. Gosain and P. N. Kapoor (1993). A comparative study for the estimators of the generalized Pareto distribution. *Journal of Hydrology*, 150, 169–185.

Nadarajah, S. (2008). Generalized Pareto models with application to drought data. *Environmetrics*, 19, 395-408.

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

Pisarenko, V. F. and D. Sornette (2003). Characterization of the frequency of extreme events by the. generalized Pareto distribution. *Pure and Applied Geophysics*, 160, 2343-2364.

R (2008), The R Project for Statistical Computing, http://www.r-project.org

Ren, F. and D.E. Giles (2010). Extreme value analysis of Canadian daily crude oil prices. *Applied Financial Economics*, 20, 941-954.

Ribatet, M. A. (2007). A user's guide to the POT package, version 1.4. http://rss.acs.unt.edu/Rdoc/library/POT/doc/guide.pdf

Ross, T. and N. Lott (2003). A climatology of 1980 – 2003 extreme weather and climate events. Technical Report 2003-01, National Climatic Data Center.

Shenton, L. R. and K. Bowman (1963). Higher moments of a maximum-likelihood estimate. *Journal of the Royal Statistical Society, B*, 25, 305-317.

Shenton, L. R. and P. A. Wallington (1962). The bias of moment estimators with an application to the negative binomial distribution. *Biometrika*, 49, 193-204.

Shi, G. H. V. Atkinson, C. M. Sellaars and C. W. Anderson (1999). Application of the generalized Pareto distribution to the estimation of the size of the maximum inclusion in clean steels. *Acta Materiala*, 47, 1455-1468.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72, 67-90.

Stansell, P. (2005). Distributions of extreme wave, crest and trough heights measured in the North Sea. *Ocean Engineering*, 32, 10156-1036.

Stephenson, A. (2008), evd: Functions for extreme value distributions, http://CRAN.R-project.org

Van Montfort, M. A. J. and J. V. Witter (1986). The generalized Pareto distribution applied to rainfall depths. *Hydrological Sciences Journal*, 31, 151-162.

Zhang, J. (2007). Likelihood moment estimation for the generalized Pareto distribution. *Australian and New Zealand Journal of Statistics*, 49, 69-77.

Zhang, J. and M. A. Stephens (2009). A new and efficient estimation method for the generalized Pareto distribution. *Technometrics*, 51, 316-325.