University of Victoria
Department of Economics

Econometrics Working Paper EWP0902

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
Victoria, B.C., Canada V8W 2Y2

&

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

Revised, April 2009

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 is found to be extremely effective in terms of bias reduction, and generally results in a small reduction in relative mean squared error. In general, the analytic bias-corrected estimators are also found to be superior to the alternative of bias-correction *via* the bootstrap.

Keywords Bias reduction; Extreme values; Generalized Pareto distribution; Peaks over threshold

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 V8W 2Y2; e-mail: dgiles@uvic.ca; Phone: (250) 721-8540; FAX: (250) 721-6214

1. Introduction

This paper discusses the calculation of analytic first-order bias expressions for the maximum likelihood estimators (MLE's) of the parameters of the generalized Pareto distribution (GPD). This distribution is widely used in extreme value analysis, and the motivation for its use 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 MLE's for the parameters of this distribution are of particular interest. Specifically, we consider the $O(n^{-1})$ bias formula introduced by Cox and Snell (1968), as re-expressed by Cordeiro and Klein (1994). This methodology is particularly appealing here, as it enables us to obtain analytic bias expressions, and hence "bias-corrected" MLE's, 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 MLE's 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, but we do not discuss this approach here. 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 in this study.

We find that bias-correcting the MLE's 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, in general this is accompanied by a modest reduction in relative mean squared error. We compare this analytic bias correction with the alternative of using the bootstrap to estimate the $O(n^{-1})$ bias, and then correcting accordingly. In common with other related studies, we find that the bootstrap bias-correction is quite ineffective, and is not to be recommended.

Section 2 summarizes the required background theory, and this is then used to derive analytic expressions for the first-order biases of the MLE's 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 MLE's. Some concluding remarks appear in section 5.

2. First-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 with respect to all derivatives up to and including the third order.

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

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

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

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

The derivatives of the cumulants are denoted:

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

and all of the 'k' expressions are assumed to be $O(n)$.

Extending earlier work by Tukey (1949), 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:

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

where k^{ij} is the $(i,j)^{\text{th}}$ 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:

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

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, …., p; and define the following matrices:

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

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

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:

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

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

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

where $\hat{K} = (K)|_{\hat{\theta}}$ and $\hat{A} = (A)|_{\hat{\theta}}$.

It can be shown that the bias of $\tilde{\theta}$ will be $O(n^{-2})$.

3. Application to the generalized Pareto distribution

We now turn to the problem of reducing the bias of the MLE's 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 (modified) scale parameter σ, are:

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

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

respectively. Note that $0 \le y < \infty$ if $\xi \ge 0$, and $0 \le y < -\sigma / \xi$ if $\xi < 0$. Our particular interest in this distribution arises in the context of modeling extreme values in the returns on financial assets.

The (integer-order) central moments of the GPD can be shown (see part A of the Appendix) 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 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 – for example, Hosking and Wallis (1987) discuss the method of moments (MOM) and probability-weighted moments estimators of θ; Castillo and Hadi (1997) propose the "elemental percentile method"; and Luceño (2006) considers various "maximum goodness of fit" estimators, based on the empirical distribution function. 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 MLE's 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 (Dugué, 1937; Cramér, 1946, p.500) do not hold if $\xi < -1/3$ (Luceño, 2006, p.905). Essentially for these reasons, maximum likelihood estimation of the parameters of the GPD can be challenging in practice, as is discussed more fully by Castillo and Hadi (1997).

Assuming independent observations, the full log-likelihood based on (12) is:

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

So,

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

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

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

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

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

$$\partial^3 l / \partial \xi^3 = 3\xi^{-4}\{2\sum_{i=1}^{n}[\ln(1+\xi y_i/\sigma)] - 2\xi\sum_{i=1}^{n}[y_i/(\sigma+\xi y_i)] - \xi^2\sum_{i=1}^{n}[y_i^{\ 2}/(\sigma+\xi y_i)^2]\} - 2(1+1/\xi)\sum_{i=1}^{n}[(y_i^3/(\sigma+\xi y_i))^3] \tag{19}$$

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

$$\partial^3 l / \partial \xi^2 \partial \sigma = -2\xi^{-2}\{\sum_{i=1}^{n}[y_i/(\sigma+\xi y_i)^2] - \sigma^{-1}\sum_{i=1}^{n}[y_i/(\sigma+\xi y_i)]\} - 2(1+1/\xi)\sum_{i=i}^{n}[y_i^{\ 2}/(\sigma+\xi y_i)^3] \tag{21}$$

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

In this case 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 MLE's 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.

As is shown in part B of the Appendix, the following results are obtained readily by direct integration after a simple change of variable:

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

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

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

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

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

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

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

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

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

Using the same change of variable, and integrating by parts, we also have:

$$E[\ln(1+\xi y/\sigma)] = \xi \,. \tag{32}$$

We can now evaluate the various terms needed to determine the Cox-Snell biases of the MLE's of ξ and σ, as discussed in section 2. Note that the $k_{ij,l}$ terms defined in (3) are not needed in the Cordeiro-Klein variant of the Cox-Snell bias formula – see (9) and the associated definitions.

The following results are easily obtained:

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

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

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

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

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

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

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

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

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

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

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

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

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

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:

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

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

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

Defining $A = [A^{(1)} \mid A^{(2)}]$, the Cox-Snell/Cordeiro-Klein expression for the biases of the MLE's of ξ and σ to order $O(n^{-1})$ is

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

which can be evaluated by using (46) – (52).

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 $\hat{B}$ can be obtained by replacing ξ and σ in (53) with their MLE's. This modified estimator is unbiased to order $O(n^{-2})$.

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 *maxLik* package (Toomet and Henningsen, 2008), primarily using the Newton-Raphson method. The use of this algorithm in similar problems is supported by Prescott and Walden (1980), for example. The Nelder-Mead algorithm was used for $n \leq 50$ in Table 1 to avoid the very occasional failure of the Newton-Raphson algorithm when the sample size is this small. Chaouche and Bacro (2006) discuss some related issues associated with solving numerically for the MLE of the GPD parameter vector. Each part of our experiment uses 50,000 Monte Carlo replications.

The results in Table 1 are *percentage* biases and MSE's, the latter being defined as $100\times$(MSE / ξ^2) and $100\times$(MSE / σ^2). These quantities are invariant to the value of the scale parameter, so we have set $\sigma = 1$. We report results for several positive values of ξ that are in the range considered by Castillo and Hadi (1997) in their simulation study of other estimators for the GPD, and are also of practical relevance. We restrict our attention to positive values for the shape parameter as (at least positive) returns on financial assets are potentially unbounded. However, other computations that we have undertaken show that the qualitative nature of our conclusions are unaltered if $\xi < 0$. The range of sample sizes that we consider is also motivated by practical applications. For example, Brooks *et al*. (2005) deal with sample sizes of $n \le 40$, while Bali and Neftci (2003) have a sample with n = 300, in each case in the context of fitting the GPD to financial data.

In Table 1 we see that the MLE's of the shape and scale parameters are negatively and positively biased, respectively. In addition, the percentage biases of the MLE's for ξ (and σ) decrease (increase) in absolute value as the true value of the shape parameter increases. Of course, these absolute biases decline monotonically as the sample size increases. The analytic bias correction performs extremely well in all cases, generally reducing the percentage biases by at least an order of magnitude. In some cases there is an "over-correction", with the percentage bias changing sign. It is encouraging that, with only one exception, the reduction in relative bias for the (corrected) estimators of ξ and σ is accompanied by a small improvement in relative mean squared error when $n \ge 50$. Similar results are reported by Cribari-Neto and Vaconcellos (2002) in the case of the beta distribution and Giles (2009) for the half-logistic distribution.

In addition to $\hat{\xi}$, $\tilde{\xi}$, $\hat{\sigma}$ and $\tilde{\sigma}$, we have also considered the bootstrap-bias-corrected estimator (Efron, 1979). The latter is obtained as $\breve{\theta} = 2\hat{\theta} - (1/N_B)[\sum_{j=1}^{N_B} \hat{\theta}_{(j)}]$, where $\hat{\theta}_{(j)}$ is the MLE of θ obtained from the j^{th} of the N_B bootstrap samples, and θ is either ξ or σ. See Efron (1982, p.33). This estimator is also unbiased to $O(n^{-2})$, but in practice it is known that this often comes at the expense of increased variance. In this part of the Monte Carlo experiment we have used 50,000 replications and 1,000 bootstrap samples within each replication – a total of 50 million evaluations for each associated entry in Table 1.

The bootstrap-corrected estimator also achieves a reasonable degree of bias reduction, especially with regard to the scale parameter. In the latter case $\%Bias(\breve{\sigma})$ is less than $\%Bias(\tilde{\sigma})$, in absolute terms, for two-thirds of the entries in Table 1. In contrast, $\%Bias(\tilde{\xi})$ is less than $\%Bias(\breve{\xi})$, in absolute terms, for two-thirds of the 21 cases considered. When estimating the shape parameter, the analytical correction results in a gain in relative bias that is at least an order of magnitude better than that of the bootstrap correction in a number of cases – *e.g.*, when n = 50 and ξ = 1.0, or n = 200 and ξ = 0.5). The percentage mean squared errors of the bootstrap-corrected estimators are generally very close to those of the analytically corrected estimators. With the exception of two cases (when n = 25), the analytical correction leads to slightly lower mean squared error than does the bootstrap correction. In addition, in a few cases (*e.g.*, for $\breve{\xi}$ when n = 150 and ξ = 1.0) the variance inflation that is induced by the bootstrap correction actually increases the percentage mean squared error above that of the original MLE. Overall, our results provide a strong case for using the Cox-Snell analytic bias correction for the MLE's of the parameters of the GPD

5. 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 $O(n^{-2})$. 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 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 improvement in relative mean squared error, at least for sample sizes of the magnitude likely to be encountered in practice. Using the bootstrap to bias-correct the maximum likelihood estimator is also quite effective for this distribution. However, on balance it is inferior to the analytic correction, especially once the effect on mean squared error is considered.

While reducing the finite-sample bias of the MLE's of the parameters of the GPD is important in its own right, there is also considerable interest in managing the bias of the MLE's of certain functions of these parameters. Specifically, in risk analysis we are concerned with value at risk (VaR) and the expected shortfall (ES), both of which 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 by the authors addresses this issue by deriving the Cox-Snell $O(n^{-1})$ biases for the estimators of VaR and ES, and evaluating the bias-corrected estimators in a manner similar to that adopted in the present paper.

Acknowledgment

We are grateful to participants at the 2009 Hawaii International Conference on Statistics, Mathematics and Related Fields, and a seminar at the University of Victoria, for their helpful comments. We also thank Lief Bluck for providing access to the additional computing facilities needed to complete the simulation experiments. The second author acknowledges financial support from King's College at the University of Western Ontario.

Table 1: Percentage biases and MSE's

n	$\%Bias(\hat{\xi})$ [$\%MSE(\hat{\xi})$]	$\%Bias(\tilde{\xi})$ [$\%MSE(\tilde{\xi})$]	$\%Bias(\breve{\xi})$ [$\%MSE(\breve{\xi})$]	$\%Bias(\hat{\sigma})$ [$\%MSE(\hat{\sigma})$]	$\%Bias(\tilde{\sigma})$ [$\%MSE(\tilde{\sigma})$]	$\%Bias(\breve{\sigma})$ [$\%MSE(\breve{\sigma})$]
			ξ = 0.5			
25	-20.1587	1.1605	-7.1104	10.0664	-8.6228	-2.6335
	[48.0904]	[82.1440]	[53.9613]	[15.2634]	[161.5268]	[14.7778]
50	-12.1930	-1.9603	-6.2334	5.9062	-1.7691	2.1070
	[25.9748]	[24.2179]	[31.4349]	[7.1023]	[10.1278]	[7.5628]
75	-5.7610	0.3024	-2.7424	3.7248	-0.5590	1.3044
	[13.3837]	[11.4417]	[20.6066]	[4.6853]	[3.5037]	[5.1182]
100	-4.2936	0.1299	-0.3207	2.7687	-0.2444	0.1146
	[9.8939]	[8.7299]	[9.6071]	[3.4162]	[2.7323]	[3.2064]
125	-4.5675	-0.9802	0.1478	2.4156	0.0035	0.3841
	[9.5717]	[8.6061]	[10.2316]	[2.7411]	[2.2631]	[2.9058]
150	-3.5011	-0.5653	-0.2740	1.9333	-0.0105	0.1147
	[7.4976]	[6.8917]	[6.2552]	[2.2364]	[1.9205]	[2.0722]
200	-2.0973	0.0538	0.8231	1.3237	-0.0673	-0.0772
	[4.7031]	[4.4580]	[4.8872]	[1.6009]	[1.4465]	[1.6480]
			ξ = 1.0			
25	-9.1557	-0.1383	0.1238	14.3801	-5.5375	-6.2403
	[19.1474]	[29.9705]	[18.0012]	[26.4306]	[142.9493]	[18.8745]
50	-4.1533	-0.0776	-0.5368	6.5035	-0.6587	-0.3049
	[8.6654]	[7.7240]	[8.4399]	[10.1828]	[7.1063]	[8.6948]
75	-2.6054	0.0780	-0.1429	4.1453	-0.3124	-0.1269
	[5.5917]	[5.2105]	[5.5304]	[6.2165]	[4.9778]	[5.6628]
100	-1.9488	0.0580	0.0515	3.0952	-0.1541	-0.0767
	[4.1462]	[3.9371]	[4.0946]	[4.5360]	[3.8564]	[4.2325]
125	-1.5604	0.0416	-0.1268	2.3881	-0.1641	0.0512
	[3.2876]	[3.1547]	[3.2764]	[3.5022]	[3.0846]	[3.2925]
150	-1.2639	0.0699	-0.0973	-1.9609	-0.1427	0.0490
	[2.7206]	[2.6302]	[2.7516]	[2.8740]	[2.5887]	[2.7528]
200	-0.9600	0.0405	0.0141	1.4940	-0.0645	0.0347
	[2.0366]	[1.9855]	[2.0268]	[2.1346]	[1.9745]	[2.0681]
			ξ = 1.5			
25	-6.0540	-0.7288	-0.2062	16.5238	-2.7281	-5.6138
	[12.3696]	[10.5364]	[12.0444]	[34.1139]	[16.5733]	[23.7049]
50	-2.7273	-0.0628	-0.3398	7.5318	-0.5637	-0.5293
	[5.8190]	[5.4434]	[5.5079]	[13.0552]	[9.4773]	[10.9063]
75	-1.7202	0.0669	-0.0591	4.8439	-0.3034	-0.2462
	[3.8004]	[3.6436]	[3.7762]	[7.9263]	[6.4518]	[7.1146]
100	-1.2862	0.0594	0.1075	3.6137	-0.1643	-0.1957
	[2.8340]	[2.7474]	[2.8040]	[5.7482]	[4.9354]	[5.3060]
125	-1.0474	0.0310	0.0519	2.8073	-0.1722	0.0143
	[2.2495]	[2.1938]	[2.2486]	[4.4312]	[3.9292]	[4.1995]
150	-0.8437	0.0562	-0.0135	2.3039	-0.1575	0.0012
	[1.8682]	[1.8304]	[1.8645]	[3.6268]	[3.2837]	[3.4491]
200	-0.6833	0.0380	0.0408	1.7519	-0.0763	0.0300
	[1.4006]	[1.3791]	[1.4004]	[2.6892]	[2.4963]	[2.5932]

References

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.

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.

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 N.J.

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.

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.

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

Giles, D. E. A. (2009), "Bias reduction for the maximum likelihood estimator of the scale parameter in the half-logistic distribution", Econometrics Working Paper EWP0901, Department of Economics, University of Victoria.

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

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.

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

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

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

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

Prescott, P. and A. T. Walden (1980), “Maximum-likelihood estimation of the parameters of the three-parameter generalized extreme-value distribution”, *Biometrika*, 67, 723-724.

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

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.

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

Toomet, O. and A. Henningsen (2008), maxLik: Maximum likelihood estimation, http://CRAN.R- project.org ; http://maxLik.org

Tukey, J. W. (1949), “Standard confidence points”, Memorandum Report No. 26, Statistical Research Group, Stanford University.

Appendix – Mathematical Derivations

A. Moments of the Generalized Pareto Distribution

A general expression for the moments of the GPD does not appear to be well documented, so it is derived here. When $\xi = 0$ the GPD collapses to the exponential distribution, whose moments are well known, so we consider only the case where $\xi \neq 0$. We will use the following result (Gradshteyn, and Ryzhik, 1965, p. 292; integral 3.241, no.4):

$$\int_0^\infty x^{\mu-1}(p+qx^\nu)^{-(m+1)}dx = (1/\nu p^{m+1})(p/q)^{\mu/\nu}\Gamma(\mu/\nu)\Gamma(1+m-\mu/\nu)/\Gamma(1+m) \qquad \text{(A.1)}$$

$$\text{(for } 0 < \mu/\nu < m+1\text{).}$$

Using the GPD density for Y in (4.2),

$$E(Y^r) = \int_0^\infty y^r \sigma^{-1}(1+\xi y/\sigma)^{-1/\xi-1}dy\,. \qquad \text{(A.2)}$$

Applying (A.1) with $\mu = r+1$; $m = 1/\xi$; $q = \xi/\sigma$; $\nu = 1$ and $p = 1$:

$$\begin{aligned} E(Y^r) &= \sigma^{-1}(\sigma/\xi)^{r+1}\Gamma(r+1)\Gamma(1+\xi-(r+1))/\Gamma(1+1/\xi) \\ &= r!\sigma^r\xi^{-(r+1)}\Gamma(1+\xi-(r+1))/\Gamma(1+1/\xi) \end{aligned}\,. \qquad \text{(A.3)}$$

Repeatedly applying the recursion formula, $\Gamma(1+a) = a\Gamma(a)$, the ratio of gamma functions in (A.3) reduces to $[(1/\xi)(1/\xi-1)(1/\xi-2).....(1/\xi-r)]^{-1}$, so that

$$\begin{aligned} E(Y^r) &= \sigma^{-1}(\sigma/\xi)^{r+1}\Gamma(r+1)\Gamma(1+\xi-(r+1))/\Gamma(1+1/\xi) \\ &= r!\sigma^r\xi^{-(r+1)}/[(1-\xi)(1-2\xi)......(1-r\xi)\xi^{r+1}] \\ &= [r!\sigma^r]/[\prod_{i=1}^{r}(1-i\xi)] \end{aligned} \qquad \text{(A.4)}$$

The parameter constraints for (A.1) to be valid amount to $0 < (r+1) < (1+1/\xi)$, or $0 < r < (1/\xi)$. This condition precisely matches the condition given by Hosking and Wallis (1987, p. 341) for the existence of the r^{th} central moment of the GPD. Finally, note that (A.4) collapses to the usual expression for the r^{th} moment of the exponential distribution when $\xi = 0$, and to that of the uniform distribution on (0 , σ) when $\xi = -1$.

B. Derivation of Equation (23)

Setting $\phi = \xi / \sigma$, and using (12) with $\xi \neq 0$,

$$E[y/(\sigma + \xi y)] = \sigma^{-1} E[y/(1+\phi y)]$$
$$= \sigma^{-2} \int_0^\infty y(1+\phi y)^{-1/\xi-2} dy$$

Let $x = 1 + \phi y$, so that $dy = dx / \phi$, and

$$E[y/(\sigma + \xi y)] = (\phi\sigma)^{-2} \int_1^\infty (x-1) x^{-1/\xi-2} dx$$
$$= (\phi\sigma)^{-2} [-\xi x^{-1/\xi} + \xi/(1+\xi) x^{-1/\xi-1}]_1^\infty$$
$$= \xi^{-2} [\xi - \xi/(1+\xi)]$$
$$= (1+\xi)^{-1}$$

Results (24) to (31) follow in a similar manner.