A new hybrid estimation method for the generalized pareto distribution

Chunlin Wang\textsuperscript{a} and Gemai Chen\textsuperscript{b,c}

\textsuperscript{a}Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada; \textsuperscript{b}Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada; \textsuperscript{c}School of Mathematics, Southwest Jiaotong University, Chengdu, Sichuan, P.R. China

\textbf{ABSTRACT} 
The generalized Pareto distribution (GPD) is important in the analysis of extreme values, especially in modeling exceedances over thresholds. Most of the existing methods for estimating the scale and shape parameters of the GPD suffer from theoretical and/or computational problems. A new hybrid estimation method is proposed in this article, which minimizes a goodness-of-fit measure and incorporates some useful likelihood information. Compared with the maximum likelihood method and other leading methods, our new hybrid estimation method retains high efficiency, reduces the estimation bias, and is computation friendly.

\textbf{1. Introduction} 
The generalized Pareto distribution (GPD) is a two-parameter distribution first introduced by Pickands (1975) having the cumulative distribution function

\[ F(x; \sigma, k) = \begin{cases} 1 - (1 - kx/\sigma)^{1/k}, & \text{if } k \neq 0, \\ 1 - \exp(-x/\sigma), & \text{if } k = 0, \end{cases} \]

and the probability density function

\[ f(x; \sigma, k) = \begin{cases} \sigma^{-1}(1 - kx/\sigma)^{1/k-1}, & \text{if } k \neq 0, \\ \sigma^{-1}\exp(-x/\sigma), & \text{if } k = 0, \end{cases} \]

where \( \sigma > 0 \) and \( k \in \mathbb{R} \) are the scale and shape parameters, respectively, and the support is \( x \in (0, \infty) \) when \( k \leq 0 \) or \( x \in (0, \sigma/k) \) when \( k > 0 \). We denote the above distribution by GPD\((\sigma, k)\).

The mean of the GPD\((\sigma, k)\) is \( \sigma/(1 + k) \) and the variance is \( \sigma^2/[(1 + k)^2(1 + 2k)], \) if they exist. In general, the \( r \)th central moment of the GPD\((\sigma, k)\) exists if \( k > -1/r \). When \( k = 1 \), the GPD\((\sigma, 1)\) becomes the uniform distribution over the interval \([0, \sigma]\); when \( k \) tends to 0, the GPD\((\sigma, k)\) converges to the exponential distribution with mean \( \sigma \); and when \( k < 0 \), the GPD\((\sigma, k)\) becomes the Pareto distribution.
Given a random sample from the GPD \((\sigma, k)\), various parameter estimation methods have been studied. As the most popular and widely used estimation method, the maximum likelihood (ML) method has been considered by Davison (1984), Smith (1984, 1985), Hosking and Wallis (1987), Grimshaw (1993), and the references cited therein. In general, the ML estimators possess high efficiency and the usual asymptotic properties when \(k < 1/2\). However, in the non-regular case (defined in Smith, 1984) where \(1/2 \leq k \leq 1\), severe convergence problems may arise according to Grimshaw (1993). When \(k > 1\), the ML estimators do not exist.

Alternatively, Hosking and Wallis (1987) studied the method of moment (MOM) estimators and the probability-weighted moment (PWM) estimators of \(\sigma\) and \(k\). The performance of the MOM estimators and the PWM estimators was evaluated together with the ML estimators in the range \(-1/2 < k < 1/2\) of the shape parameter. They concluded that the MOM was unreliable for \(k < -0.2\), the PWM method performed well only when \(-1/2 < k < 0\), and the ML method needed a sample size as large as \(n = 500\) to reach its asymptotic efficiency. Additionally, when \(k \leq -1/2\), the MOM estimators do not exist, and when \(k \leq -1\) the PWM estimators do not exist. It was noted that even when both of the MOM and PWM estimators exist, they may produce unacceptable results because some of the sample points may fall outside the estimated endpoint \(\hat{\sigma}/\hat{k}\). The infeasibility of the PWM estimators was also identified in Chen and Balakrishnan (1995). To remedy this problem, Dupuis and Tsao (1998) derived a simple constraint on feasibility into the MOM and PWM estimators. However, all of the estimators based on moments are known to have low large-sample efficiencies and are available within a very restricted region of the parameter space.

Castillo and Hadi (1997) proposed an elemental percentile method (EPM) which is assumed to work for all possible values of the parameters. The idea is to make use of the order statistics by initially equating GPD \((\sigma, k)\) to all pairs of the order statistics, and to use the medians as the final estimators of \(\sigma\) and \(k\). Compared with the MOM and the PWM methods, their simulation results indicated that there was no dominant method in terms of bias and root mean squared error (MSE) when \(k\) is restricted in \(-2 \leq k \leq 2\).

Luceño (2006) investigated a maximum goodness-of-fit (MGF) method based on the family of empirical distribution function (EDF) statistics. By minimizing any of the EDF statistics measuring the discrepancy between \(F(x; \sigma, k)\) given in (1) and the EDF with respect to the unknown parameters \(\sigma\) and \(k\), this method can always yield valid estimates over the whole parameter space. The MGF method when the Anderson–Darling statistic is used gives the most balanced performance. However, the bivariate search for the minimum could be slow and convergence problems arise if without a well-specified initial point.

Zhang and Stephens (2009) and Zhang (2010) introduced a new efficient estimation method based on the likelihood and the empirical Bayes method (EBM). This method is computationally friendly and the data-driven prior can ensure that the parameter estimates are valid. However, the performance of the EBM estimators turned out to be dependent on a careful choice of the prior distribution. In Zhang and Stephens (2009), the prior distribution was chosen as the GPD itself with the shape parameter \(\tilde{k} = -1/2\). To improve the poor performance of the EBM in the heavy-tailed situation \((k < -1)\), a modified EBM, denoted by EBM*, was introduced in Zhang (2010). The EBM* was found to dominate the other existing competitors in the extended range \(-6 \leq k \leq 1/2\) in terms of estimation bias and efficiency.

The goal of this article is to develop a new hybrid parameter estimation method for the GPD to improve upon the existing methods. In Section 2, we provide the motivation and the details of the new hybrid method. In Section 3, the performance of the proposed hybrid
method is evaluated against several significant competitors via simulation. In Section 4, we illustrate our new hybrid method through revisiting the Bilbao sea waves data analyzed in the literature, and in Section 5 we draw some conclusions.

2. A new hybrid parameter estimation method

The basis of our new hybrid method is the MGF method when the Anderson–Darling statistic is used. Given a random sample \(X = \{X_1, \ldots, X_n\}\) from the GPD \((\sigma, k)\) with distribution function \(F(x; \sigma, k)\) given in (1) and define the indicator function \(I_{X_i}(x) = 1\) if \(X_i \leq x\) and \(I_{X_i}(x) = 0\) if \(X_i > x\). The EDF of the sample is given by

\[
F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{X_i}(x).
\]

The Anderson–Darling statistic \(A^2(\sigma, k)\) is defined as

\[
A^2(\sigma, k) = n \int_{-\infty}^{\infty} \{F_n(x) - F(x; \sigma, k)\}^2 \{F(x; \sigma, k)(1 - F(x; \sigma, k))\}^{-1} dF(x; \sigma, k).
\]

Denoting the \(i\)th order statistic of the sample by \(X_{(i)}\) and applying the probability integral transformation to get \(Z_i = F(X_{(i)}; \sigma, k), i = 1, \ldots, n\), we can express the Anderson–Darling statistic \(A^2(\sigma, k)\) as

\[
A^2(\sigma, k) = -n - \frac{1}{n} \sum_{i=1}^{n} \left\{ (2i - 1) \log Z_i + (2n + 1 - 2i) \log(1 - Z_i) \right\}.
\]

The MGF method estimators \(\hat{\sigma}_{MGF}\) and \(\hat{k}_{MGF}\) are obtained by minimizing the above Anderson–Darling statistic with respect to the unknown parameters \(\sigma\) and \(k\) over the parameter space \(\mathcal{A} = \{(\sigma, k) \in \mathbb{R}^2 : k \leq 0, \sigma > 0\} \cup \{k > 0, \sigma/k > X_{(n)}\}\). This method uses the information from \(F(x; \sigma, k)\) and all order statistics. It ensures that the parameter estimators are valid and the estimators obtained have small biases, but it suffers from a slow bivariate search for minimum and more seriously from convergence problems when \(k \leq -1\) (the final estimates are very sensitive to the initial values \((\sigma^{(0)}, k^{(0)})\)). The MGF estimators also have low efficiency compared with the ML estimators if exist.

The log-likelihood function for the parameters \(\sigma\) and \(k\) in GPD \((\sigma, k)\) is given by

\[
\ell(\sigma, k) = -n \log \sigma - \left(1 - \frac{1}{k}\right) \sum_{i=1}^{n} \log \left(1 - \frac{kX_i}{\sigma}\right),
\]

and the score equations for estimating \(\sigma\) and \(k\) are

\[
n(k-1) = \sum_{i=1}^{n} \log \left(1 - \frac{kX_i}{\sigma}\right) + (k-1) \sum_{i=1}^{n} \left(1 - \frac{kX_i}{\sigma}\right)^{-1},
\]

\[
k = -\frac{1}{n} \sum_{i=1}^{n} \log \left(1 - \frac{kX_i}{\sigma}\right).
\]

As pointed out by Davison (1984), if we define \(\theta = k/\sigma\) and reparameterize \((\sigma, k)\) into \((\theta, k)\), we can use the second score equation to link \(k\) to \(\theta\) through the relation
\( k = -n^{-1} \sum_{i=1}^{n} \log (1 - \theta X_i) \), and obtain the profile log-likelihood function for \( \theta \) as

\[
\ell(\theta) = -n - \sum_{i=1}^{n} \log (1 - \theta X_i) - n \log \left[ -\frac{1}{n\theta} \sum_{i=1}^{n} \log (1 - \theta X_i) \right].
\]

Suppose that a local maximum of \( \ell(\theta) \) can be found at \( \hat{\theta}_{\text{MLE}} \) numerically over the space \( \mathcal{B} = \{ \theta \in \mathbb{R} : \theta < 1/X(n) \} \), then the corresponding ML estimators \( \hat{\sigma}_{\text{MLE}} \) and \( \hat{k}_{\text{MLE}} \) of \( \sigma \) and \( k \) are given by

\[
\hat{k}_{\text{MLE}} = -n^{-1} \sum_{i=1}^{n} \log (1 - \hat{\theta}_{\text{MLE}} X_i) \quad \text{and} \quad \hat{\sigma}_{\text{MLE}} = \frac{\hat{k}_{\text{MLE}}}{\hat{\theta}_{\text{MLE}}}.
\]

However, maximizing \( \ell(\theta) \) can still be problematic. As shown in Grimshaw (1993), there can be more than one root to \( d\ell(\theta)/d\theta = 0 \), and a carefully chosen initial value \( \theta^{(0)} \) is essential. An algorithm for computing the ML estimators based on \( \ell(\theta) \) was designed in Grimshaw (1993). The interested reader can consult the supplementary materials for this article where proportions of the convergence problem that occurred for both the ML method and the MGF method are reported, based on the algorithm of Grimshaw (1993) and the \texttt{R} function \texttt{gpdmgf} available in package \texttt{POT}, respectively.

The MGF estimators enjoy small estimation bias and wide feasibility, while the high estimation efficiency of the ML estimators is also desired. We, however, want to avoid the computational issues associated with these two methods. These considerations motivate us to study the following hybrid parameter estimation method, which may retain a high degree of efficiency as well as small bias, and is computation friendly.

Under the parameterization of \((\theta, k)\) for the GPD and utilizing the likelihood-based link \( k = -n^{-1} \sum_{j=1}^{n} \log (1 - \theta X_j) \), we estimate \( \theta \) through

\[
\hat{\theta}_{\text{HYB}} : = \arg \min_{\theta \in \mathcal{B}} G(\theta) = \arg \min_{\theta \in \mathcal{B}} A^2(\theta, k), \tag{2}
\]

where \( G(\theta) = A^2(\theta, k) \) and \( \mathcal{B} = \{ \theta : \theta < 1/X(n) \} \). Let

\[
g_i(\theta) = 1 - (1 - \theta X_{(i)})^{-n/\sum_{j=1}^{n} \log (1 - \theta X_j)}, \quad i = 1, \ldots, n,
\]

then the computation formula for the target function \( G(\theta) \) is

\[
G(\theta) = -n - \frac{1}{n} \sum_{i=1}^{n} \left\{(2i - 1) \log[g_i(\theta)] + (2n + 1 - 2i) \log[1 - g_i(\theta)]\right\}.
\]

It is much more convenient to minimize this function \( G(\theta) \) in one dimension than the ML and MGF methods subject to \( \theta < 1/X(n) \) to obtain our hybrid estimator \( \hat{\theta}_{\text{HYB}} \) of \( \theta \). Then our hybrid estimators \( \hat{k}_{\text{HYB}} \) and \( \hat{\sigma}_{\text{HYB}} \) of \( k \) and \( \sigma \) can be found from

\[
\hat{k}_{\text{HYB}} = -\frac{1}{n} \sum_{i=1}^{n} \log (1 - \hat{\theta}_{\text{HYB}} X_i) \quad \text{and} \quad \hat{\sigma}_{\text{HYB}} = \frac{\hat{k}_{\text{HYB}}}{\hat{\theta}_{\text{HYB}}}.
\]

It seems not easy to derive the (asymptotic) variances for the hybrid estimators due to the hybrid mechanism. However, the bootstrap method introduced by Efron (1979) can provide useful approximations to the distributions of the hybrid estimators. Using the bootstrap method to find standard errors for other estimators of GPD has already been done by many authors, such as Castillo and Hadi (1997) for their EPM estimators and Zhang and Stephens (2009) for their EBM estimators. However, the potential inconsistency of the parametric bootstrap distributions due to the irregularity of the GPD leads us to examine several commonly
used bootstrap procedures via simulation in the next section. We have written an R function to compute the suggested bootstrap standard errors and confidence intervals for the hybrid estimators, which is available upon request.

3. Finite-sample performance comparison

Based on the performance comparison results for various estimators in the literature, we only select the top performers in our performance comparison. This includes the ML estimators, the MGF estimators, and the EBM* estimators.

Data are generated from \( F(x; \sigma, k) \) for \( \sigma = 1 \) and \(-6 \leq k \leq 2\). There is no loss of generality to fix \( \sigma \) at 1 because of its invariance property. This parameter range of \( k \) contains all published ranges in the literature as subsets and is also of theoretical and practical concerns. First, this range of \( k \) covers the exponential distribution when \( k = 0 \) and the uniform distribution when \( k = 1 \). Second, many real-world examples show the possibility of non regular case where estimates of \( k > 1/2 \), such as the two examples in Castillo and Hadi (1997), and one of them will be revisited in the next section. Third, estimates of \( k \leq -1 \) are often observed in heavy-tailed financial data.

Since the ML method can lead to severe convergence problems, we follow Luceño (2006) to use the quasi-maximum likelihood (QML) estimates

\[
\hat{k}_{\text{QML}} = -\frac{1}{n-1} \sum_{i=1}^{n-1} \log \left( 1 - \frac{X(i)}{X(n)} \right) \quad \text{and} \quad \hat{\sigma}_{\text{QML}} = \hat{k}_{\text{QML}} X(n)
\]

as the ML estimates when the ML method fails to converge, that is, when the algorithm of Grimshaw (1993) returns “MLE does not converge.” It is well known that the ML method does not have a finite estimate of \( k \) when \( k > 1 \). To accommodate this boundary condition issue, we also use the QML estimates if the computed \( \hat{k} > 1 \) in our simulation.

We consider sample sizes \( n = 20, 50, 100, \) and 200. For a given sample size \( n \) and a given value of \( k \) in the interval \([-6, 2]\), we simulate 10,000 samples of size \( n \) from \( F(x; 1, k) \) and estimate \( \sigma \) and \( k \) using our hybrid method and the other methods. For each method, we compute the bias and the MSE for each parameter estimator. For sample sizes \( n = 50 \) and \( n = 200 \), the biases for the various estimators of \( \sigma \) and \( k \) are plotted against \( k \) in Fig. 1, and the MSEs are plotted against \( k \) in Fig. 2.

We note that the MGF method of Luceño (2006) was not compared with the EBM* method in Zhang (2010) or in other published literature. Luceño (2006) used the root mean squared error (RMSE), while Zhang (2010) used the relative efficiency (the MSE divided by the Cramér–Rao lower bound) as their evaluation measures. In our simulation studies, we found that the RMSEs were usually too close to each other to tell the differences graphically. On the other hand, the Cramér–Rao lower bound does not exist for \( k > 1/2 \). We therefore choose the MSE to compare the different estimators.

From Fig. 1, we see that our hybrid method generally outperforms the other three methods in terms of estimation bias. In terms of MSE, we see from Fig. 2 that our hybrid method outperforms the ML method (especially in estimating \( \sigma \)) and the MGF method (especially in estimating \( k \)) and performs better or the same as the EBM* method for \(-6 \leq k < 0\), slightly worse than the EBM* method for \( 0 \leq k < 1 \) and surprisingly better than the EBM* method for \( 1 \leq k \leq 2 \). The poor performance of the popular EBM* method for \( 1 \leq k \leq 2 \) is probably due to the fact that the data-driven prior used in EBM* was set up when \(-6 \leq k \leq -1 \) was well looked after, but no consideration was given to the case of \( k > 1/2 \) because in this case
the performance measure (relative efficiency) is not defined. In comparison, our proposed hybrid method behaves well in such a non regular range of $k$.

Since we would expect asymmetry in the sampling distributions for $\hat{k}$ and $\hat{\sigma}$ due to perhaps the heavy-tail behavior and the boundary condition, it is worth comparing the results for several bootstrap procedures. For the proposed hybrid estimators, we carry out a simulation study comparing three bootstrap percentile confidence intervals based on the parametric bootstrap, the non parametric bootstrap, and the skew-corrected and accelerated bootstrap ($BC_a$) described in Efron (1987) and implemented in R package boot. In the simulation, we consider 2,000 replications with 1,000 bootstrap samples for each replication. The scale parameter $\sigma$ is fixed at 1 as before and $k$ values are selected between $-6 \leq k \leq 2$. Table 1 reports the coverage probabilities and the average lengths (in parentheses) for the three types of 95% bootstrap percentile confidence intervals when the sample size $n = 50$. Similar conclusions can be drawn for 90% confidence level and other sample sizes.

It can be seen from Table 1 that in general all three bootstrap percentile confidence intervals perform well, except for $k = 1$ (uniform distribution over $[0, \sigma]$) and $k = 2$. For both

Figure 1. Bias comparisons for estimating the GPD parameters $\sigma$ and $k$, based on 10,000 replications with sample size $n = 50$ and $n = 200$. 
Figure 2. MSE comparisons for estimating the GPD parameters $\sigma$ and $k$, based on 10,000 replications with sample size $n = 50$ and $n = 200$.

Table 1. The simulated coverage probabilities and average lengths (in parentheses) of three types of bootstrap percentile confidence intervals for $\hat{k}_{\text{HYB}}$ and $\hat{\sigma}_{\text{HYB}}$.

| $k$  | Coverage probabilities (lengths) for $\hat{\sigma}_{\text{HYB}}$ | Coverage probabilities (lengths) for $\hat{k}_{\text{HYB}}$ |
|------|---------------------------------------------------------------|---------------------------------------------------------------|
|      | Parametric bootstrap  | Non parametric bootstrap  | BC$_{\alpha}$ | Parametric bootstrap  | Non parametric bootstrap  | BC$_{\alpha}$ |
| $-6.0$ | 0.9490 (3.1386)  | 0.9515 (3.4075)  | 0.9485 (3.2155) | 0.9460 (3.9048)  | 0.9350 (3.8338)  | 0.9465 (4.0301) |
| $-3.0$ | 0.9505 (1.8843)  | 0.9505 (1.9491)  | 0.9480 (1.8910) | 0.9435 (2.2189)  | 0.9325 (2.1883)  | 0.9395 (2.2881) |
| $-1.0$ | 0.9465 (1.1870)  | 0.9435 (1.9970)  | 0.9440 (1.8144) | 0.9310 (1.1475)  | 0.9260 (1.1472)  | 0.9270 (1.1832) |
| 0.0   | 0.9555 (0.8860)  | 0.9520 (0.8984)  | 0.9535 (0.8797) | 0.9615 (0.7373)  | 0.9465 (0.7660)  | 0.9505 (0.7831) |
| 0.5   | 0.9530 (0.7601)  | 0.9535 (0.7880)  | 0.9570 (0.7903) | 0.9520 (0.6884)  | 0.9475 (0.7480)  | 0.9580 (0.7582) |
| 1.0   | 0.6270 (0.7229)  | 0.9425 (0.8145)  | 0.8475 (0.7872) | 0.6260 (0.9886)  | 0.9475 (1.0926)  | 0.8455 (1.0651) |
| 2.0   | 0.9185 (0.6284)  | 0.9495 (0.6692)  | 0.9535 (0.6750) | 0.9130 (1.2755)  | 0.9490 (1.3804)  | 0.9525 (1.3879) |

The confidence level is 95%, the sample size $n = 50$, and $\sigma = 1$. 
Table 2. The estimated GPD parameters for the Bilbao waves exceedances at 6 given thresholds using various estimation methods.

| \( t \) | \( \hat{\sigma} \) | \( \hat{k} \) |
|-------|-------|-------|
|       | MOM   | PWM   | MLE   | EBM*  | MGF   | HYB   | MOM   | PWM   | MLE   | EBM*  | MGF   | HYB   |
| 7.0   | 2.748 | 2.778 | 2.501 | 2.331 | 2.451 | 2.449 | 1.052 | 1.074 | 0.861 | 0.782 | 0.838 | 0.839 |
| 7.5   | 1.622 | 1.618 | 1.860 | 1.722 | 1.632 | 1.630 | 0.606 | 0.602 | 0.768 | 0.686 | 0.614 | 0.623 |
| 8.0   | 1.385 | 1.371 | 1.647 | 1.462 | 1.417 | 1.416 | 0.647 | 0.630 | 0.864 | 0.751 | 0.682 | 0.693 |
| 8.5   | 1.130 | 1.115 | 1.400 | 1.146 | 1.176 | 1.175 | 0.722 | 0.700 | 1.000 | 0.767 | 0.789 | 0.799 |
| 9.0   | 0.814 | 0.809 | 0.900 | 0.756 | 0.846 | 0.845 | 0.833 | 0.823 | 1.000 | 0.760 | 0.900 | 0.908 |
| 9.5   | 0.626 | 0.601 | 0.400 | 0.361 | 0.521 | 0.515 | 1.709 | 1.601 | 1.000 | 0.736 | 1.291 | 1.279 |

the parametric bootstrap and the BC\(_a\), when \( k = 1 \), their coverage probabilities are quite a bit lower than the 95% confidence level. Since we know that the standard \( n \)-out-of-\( n \) parametric bootstrap has been proved to be inconsistent when the ML estimators are used in this case, the parametric bootstrap and the BC\(_a\) based on our hybrid estimation may suffer the same way. When \( k = 2 \), the parametric bootstrap also produces slightly lower coverage probabilities (under 92%). On the other hand, the non parametric bootstrap has stable coverage probabilities around the 95% level with sometimes slightly longer interval lengths for the parameter range considered. Therefore, we recommend the non parametric bootstrap procedure for making inference about the hybrid estimators in practice.

4. Real data example

To illustrate the use of our new hybrid estimation method, a real-world example originally analyzed in Castillo and Hadi (1997) is revisited here. The dataset consists of the zero-crossing hourly mean periods (in seconds) of the sea waves measured in a Bilbao buoy, Spain, in January, 1997. Later on, this dataset was analyzed in Lučenko (2006) and in Zhang and Stephens (2009). Only the 197 observations with periods above 7 s are taken into consideration. Following the above-mentioned authors, we fit this dataset to GPD(\( \sigma, k \)) using thresholds at \( t = 7.0, 7.5, 8.0, 8.5, 9.0, \) and 9.5. In Table 2, the parameters \( \sigma \) and \( k \) are estimated using different estimation methods (see also Table 4 in Castillo and Hadi, 1997, Fig. 3 in Lučenko, 2006 and Table 3 in Zhang and Stephens, 2009). Note that because the ML estimators do not exist if \( k > 1 \), instead of using NAs, we use the estimators \( \hat{k} = 1 \) and \( \hat{\sigma} = \max X_i \) if the computed \( \hat{k} > 1 \) in Table 2.

We can observe from Table 2 that all the methods give estimates \( \hat{k} \) outside the commonly used range \(-1 < k < 1/2\). In this case, the MOM and the PWM estimates are deficient and may give infeasible results, and the ML estimates are computationally unstable especially for the small sample sizes, and thus is disfavored.

Using the parameter estimation results at threshold \( t = 7.5 \), we plot in Fig. 3 the GPD function fitted by the new hybrid estimators together with the EDF and also the quantile-quantile plot for the Bilbao waves data (see also Fig. 2 in Lučenko, 2006, and Fig. 4 in Zhang and Stephens, 2009). We see from Fig. 3 that the new hybrid method gives an overall good fit to the Bilbao waves data.

The non parametric bootstrap standard errors for \( \hat{\sigma}_{HYB} \) and \( \hat{k}_{HYB} \) are found to be \( se(\hat{\sigma}_{HYB}) = 0.186 \) and \( se(\hat{k}_{HYB}) = 0.103 \). The corresponding 95% non parametric bootstrap percentile confidence intervals for \( \sigma \) and \( k \) are (1.282, 2.024) and (0.409, 0.823), respectively. The bootstrap confidence interval for \( k \) implies that \( k \) is significantly different from 0 and 1 and is likely to belong to the non regular case.
Figure 3. Bilbao waves data: the left panel is plot of the GPD function fitted by the hybrid estimators versus the empirical distribution function; and the right panel is the quantile–quantile plot. The plots are both for the threshold $t = 7.5$.

5. Conclusions

In this article, we have developed a new hybrid parameter estimation method for the GPD. Unlike the MOMs, the PWM method, and the ML method, the new hybrid method can always provide valid estimates for a wide parameter space investigated. Unlike the ML method and the MGF method, the new hybrid method does not suffer from the convergence problem and is easier and faster to implement. Compared with the top performer proposed by Zhang (2010), the new hybrid method offers smaller estimation bias and comparable or smaller mean squared error. Coupled with the non parametric bootstrap procedure, the new hybrid method can allow practitioners to conduct reliable and accurate data analysis using the GPD.

Acknowledgement

Gemai Chen’s research was supported by a grant from Natural Sciences and Engineering Research Council of Canada.

References

Castillo, E., Hadi, A., (1997). Fitting the generalized Pareto distribution to data. J. Am. Stat. Assoc. 92:1609–1620.
Chen, G., Balakrishnan, N. (1995). The infeasibility of probability weighted moments estimation of some generalized distribution. In: Balakrishnan, N., ed. Recent Advances in Life-Testing and Reliability (pp. 565–573). London: CRC Press.
Davison, A. (1984). Modelling excesses over high thresholds, with an application. In: de Oliveira, J.T., ed. Statistical Extremes and Applications (pp. 461–482). Dordrecht: Riedel.
Dupuis, D., Tsao, M. (1998). A hybrid estimator for generalized Pareto and extreme-value distributions. Commun. Stat.-Theory Methods 27(4): 925–941.
Efron, B. (1979). Bootstrap methods: another look at the jackknife. Ann. Stat. 7(1): 1–26.
Efron, B. (1987). Better bootstrap confidence intervals. J. Am. Stat. Assoc. 82(397): 171–185.
Grimshaw, S. (1993). Computing maximum likelihood estimates for the generalized Pareto distribution. Technometrics 35(2): 185–191.
Hosking, J., Wallis, J. (1987). Parameter and quantile estimation for the generalized Pareto distribution. Technometrics 29(3): 339–349.
Luceño, A. (2006). Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators. Comput. Stat. Data Anal. 51(2): 904–917.
Pickands, J. (1975). Statistical inference using extreme order statistics. Ann. Stat. 3(1): 119–131.
Smith, R.L. (1984). Threshold methods for sample extremes. In: de Oliveira, J.T., ed. Statistical Extremes and Applications (pp. 621–638). Dordrecht: Riedel.
Smith, R.L. (1985). Maximum likelihood estimation in a class of nonregular cases. Biometrika 72(1): 67–90.
Zhang, J. (2010). Improving on estimation for the generalized Pareto distribution. Technometrics 52(3): 335–339.
Zhang, J., Stephens, M. (2009). A new and efficient estimation method for the generalized Pareto distribution. Technometrics 51(3): 316–325.