A Robust Estimator for the Tail Index of Pareto-type Distributions

B. Vandewalle [a,b], J. Beirlant [c,*], A. Christmann [d], M. Hubert [c]

[a] *Department of Mathematics, Katholieke Universiteit Leuven, Belgium*

[b] *Center for Statistics and Applications, Universidade de Lisboa, Portugal*

[c] *Department of Mathematics and University Center of Statistics, Katholieke Universiteit Leuven, Belgium*

[d] *Department of Mathematics, Vrije Universiteit Brussel, Belgium*

Abstract

In extreme value statistics, the extreme value index is a well-known parameter to measure the tail heaviness of a distribution. Pareto-type distributions, with strictly positive extreme value index (or tail index) are considered. The most prominent extreme value methods are constructed on efficient maximum likelihood estimators based on specific parametric models which are fitted to excesses over large thresholds. Maximum likelihood estimators however are often not very robust, which makes them sensitive to few particular observations. Even in extreme value statistics, where the most extreme data usually receive most attention, this can constitute a serious problem. The problem is illustrated on a real data set from geopedology, in which a few abnormal soil measurements highly influence the estimates of the tail index. In order to overcome this problem, a robust estimator of the tail index is proposed, by combining a refinement of the Pareto approximation for the conditional distribution of relative excesses over a large threshold with an integrated squared error approach on partial density component estimation. It is shown that the influence function of this newly proposed estimator is bounded and through several simulations it is illustrated that it performs reasonably well at contaminated as well as uncontaminated data.

Key words: Extreme value statistics; relative excesses over a large threshold; robust tail index estimation; geopedology.

---

* Correspondence to: J. Beirlant, Department of Mathematics and University Center of Statistics, Katholieke Universiteit Leuven, W. de Croylaan 54, 3000-Leuven, Belgium

Email address: `Jan.Beirlant@wis.kuleuven.be` (J. Beirlant).

Preprint submitted to Elsevier Science *January 8, 2007*

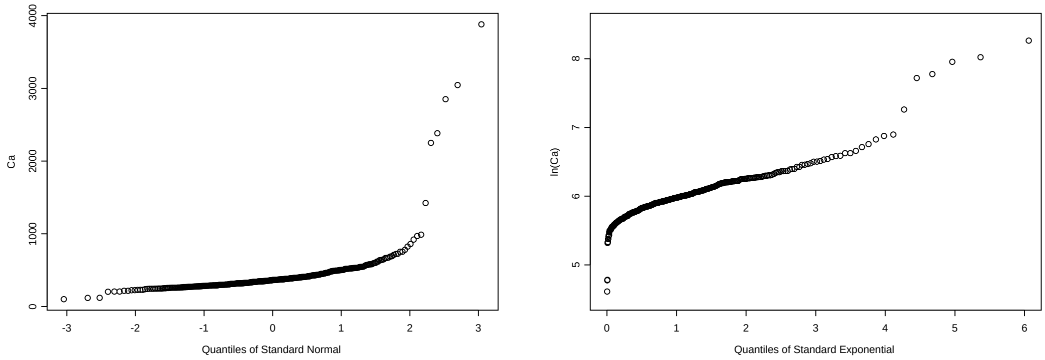
1 INTRODUCTION

In extreme value statistics emphasis lies on the modelling of rare events, mostly events with a low frequency but a high impact. Common practice is to characterize the size and frequency of such extreme events mainly by the extreme value index γ. The most prominent estimators of this real-valued parameter are maximum likelihood estimators of specific parametric models which are fitted to excesses over large thresholds (see Smith, 1987). These maximum likelihood estimators often prove to be highly efficient, though non-robust against deviations of the actual distribution from the assumed parametric model. This is for instance the case in the presence of outliers or suspicious data, where the performance of the maximum likelihood estimators and the quality of the corresponding estimates of the tail index are often seriously affected.

As shown in Brazauskas and Serfling (2000a), small errors in the estimation of the tail index can already produce large errors in the estimation of quantities based on the tail index γ. Hence, some robust algorithms replacing the maximum likelihood part of this methodology, can be of use, leading to an alternative set of outlier-resistant estimators of the tail index γ and quantities derived from it such as extreme quantiles, small exceedance probabilities and mean excess functions, which are often used in hydrology, meteorology, pedology, structural engineering, economics and actuarial science (e.g. as principles on which to set reinsurance premiums).

In this paper, we study soil data from the Condroz region in Belgium, gathered by the Unit of Geopedology at Gembloux Agricultural University. For more information, see Laroche and Oger (1999) and Goegebeur et al. (2005). In agriculture, a new concept of crop management has emerged, permitting within-field variation of crop techniques as for instance the adjustment of fertilizer inputs on the basis of soil sampling and analysis. The development of this precision farming has drastically increased the demand for soil data and laboratories are now burdened with large data sets, inevitably causing concern about outliers and their influence on the quality of the information. Therefore, robust estimation procedures have become a necessity in the database management of soil data in order to provide high quality information.

Our analysis is limited to Calcium (Ca) records from one of the communities in the Condroz region (428 observations, see Goegebeur et al., 2005). The Ca distribution at higher pH-levels appears to be right-skewed and long-tailed, resulting in rather frequent large Ca measurements, as can be seen in the normal quantile plot of Ca given in Figure 1(a). As a result, robust statistical procedures which assume that the regular data points are sampled from a normal distribution will flag too many large observations as outliers. Such long tailed data can be analyzed more efficiently in the framework of extreme value theory. Nevertheless, as will be seen later on, even in the background of a heavy tailed model some observations will still appear to be suspicious.



(a) (b)

Figure 1. (a) Normal quantile plot and (b) Pareto quantile plot of Ca for one of the communities of the Condroz region in Belgium.

Robust methods for extreme values have already been discussed in recent literature. For example, Brazauskas and Serfling (2000b) consider robust estimation in the context of strict Pareto distributions. Dupuis and Field (1998), respectively Peng and Welsh (2001) and Juárez and Schucany (2004), derived robust estimation methods for the case where the observations follow a generalized extreme value distribution, respectively a generalized Pareto distribution, light or heavy tailed. In this paper, we propose a new robust tail index estimation procedure for the semi-parametric setting of Pareto-type (or heavy-tailed) distributions. So here, the strict Pareto distribution is assumed to hold only asymptotically, i.e. for excess distributions over high enough threshold values.

In Section 2 we give a brief introduction to univariate extreme value statistics and some well-known estimators of the tail index γ for Pareto-type distributions. In Section 3 a new estimator for γ is introduced, based on a recently developed refinement of the Pareto approximation for the conditional distribution of relative excesses over a large threshold (Beirlant et al., 2004b) in combination with an integrated squared error approach on partial density component estimation (Scott, 2004). We investigate its robustness properties in Section 4 by means of the influence function, in the context of heavy tailed models contaminated with outliers. Throughout, the discussed methods are applied to the Condroz data and in Section 5 their finite-sample behavior is investigated, both for contaminated and uncontaminated simulated data sets.

2 EXTREME VALUE STATISTICS

In this section, we give a brief introduction to some well-known estimators of the extreme value index for Pareto-type distributions, i.e. with positive tail index γ.

2.1 Extreme value theory

Consider $X_1, \ldots, X_n$ independent and identically distributed (i.i.d.) random variables with common distribution function F and quantile function Q. Denote the corresponding order statistics by $X_{1,n} \leq \ldots \leq X_{n,n}$ and suppose there exist sequences of constants $(a_n > 0)$ and $(b_n \in \mathbb{R})$ such that the properly centered and normed sample maxima $X_{n,n}$ converge in distribution to a non-degenerate limit H, i.e.

$$\lim_{n\to\infty} P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = H(x),$$

for all continuity points of H. Then, the limit distribution H is necessarily of generalized extreme value type (Fisher and Tippett, 1928),

$$H_\gamma(x) = \begin{cases} \exp\Big(- (1+\gamma x)^{-\frac{1}{\gamma}} \Big), & 1+\gamma x > 0, \gamma \neq 0, \\ \exp\Big(- \exp(-x) \Big), & x \in \mathbb{R}, \gamma = 0, \end{cases}$$

and F is said to belong to the maximum domain of attraction of H_γ, denoted as $F \in \mathcal{D}(H_\gamma)$. Most common continuous distribution functions satisfy this weak condition quite naturally.

The real-valued parameter γ is referred to as the tail index or extreme value index of F. Distributions for which $\gamma > 0$ are called Pareto-type (or heavy tailed) distributions, as the tail typically decays polynomially, i.e. $\bar{F}(x) = x^{-1/\gamma} l_F(x)$ with l_F a slowly varying function. Examples in this class are the Fréchet, Pareto and Burr distributions. The Gumbel class of distributions with $\gamma = 0$ is a quite extensive class of distributions, mainly with exponentially decreasing tails. Examples are the Exponential, Normal and Gamma distributions. The Weibull class, with $\gamma < 0$ consists of distributions with a finite right endpoint x_+ for which $\bar{F}(x_+ - 1/x) = x^{1/\gamma} l_F(x)$ with l_F again a slowly varying function. Here, examples are the Uniform, Beta and reversed Burr distributions. For a general reference on extreme value theory we refer for instance to Beirlant et al. (2004a). In this paper, we will concentrate on Pareto-type distributions, i.e. distributions for which there exists a $\gamma > 0$ such that

$$\bar{F}(x) = x^{-1/\gamma} l_F(x), \quad \text{or equivalently,} \quad U(x) = x^{\gamma} l_U(x), \tag{1}$$

where $U(x) = Q(1-1/x)$ and l_F and l_U are slowly varying functions at infinity. Here, a function l is slowly varying at infinity when $l(tx)/l(t) \to 1$, as $t \to \infty$, for all $x > 0$. The main focus here is on the estimation of the tail index γ in the background of a Pareto-type tail.

2.2 Tail index estimation

Originally, the popular Hill (1975) estimator was introduced as a maximum likelihood estimator. In the Pareto-type case, the conditional distribution F_{Y_t} of relative excesses $Y_t = \left(\frac{X}{t} \;\middle|\; X > t\right)$ over a threshold t satisfies

$$\bar{F}_{Y_t}(y) = P\left(\frac{X}{t} > y \;\middle|\; X > t\right) = y^{-1/\gamma} \frac{l_F(ty)}{l_F(t)}, \tag{2}$$

for all $y \geq 1$ and l_F as in expression (1). Since, for slowly varying functions, $l_F(ty)/l_F(t) \to 1$ for all $y > 0$, it is easily seen that ultimately

$$F_{Y_t}(y) \to 1 - y^{-1/\gamma}, \qquad \text{as } t \to \infty. \tag{3}$$

Assuming the above Pareto approximation to hold exactly as a model for the conditional distribution of the relative excesses $Y_{jk} = X_{n-j+1,n}/t$ $(j = 1, \ldots, k)$ above a high threshold $t = X_{n-k,n}$ leads to the maximum likelihood estimator

$$\hat{\gamma}_{k,H} = \frac{1}{k} \sum_{j=1}^{k} \log\left(X_{n-j+1,n}/X_{n-k,n}\right).$$

This estimator can also easily be considered as a slope estimator for the linear part of the Pareto quantile plot

$$\left(\log\left(\frac{n+1}{j}\right), \log X_{n-j+1,n}\right), \qquad j = 1, \ldots, n,$$

as it measures the average increase in the plot above a certain anchor point $\left(\log(\frac{n+1}{k+1}), \log X_{n-k,n}\right)$. For this, we can refer for instance to Kratz and Resnick (1996) and Beirlant et al. (1996).

However, using the maximum likelihood estimator point of view, the assumption that for a Pareto-type distribution, above a certain threshold, the relative excesses behave as (ordered) data from a strict Pareto distribution is sometimes over-optimistic. This mostly happens when the slowly varying part of expression (2) disappears at a very slow rate, in many instances resulting in severe bias. In some applications, the resulting - and often substantial - bias that appears with this estimator, even when the model assumption of Pareto-type behavior is met, can constitute a serious problem. Systematic over- or underestimation of the tail index not only produces large errors in the estimation of quantities based on the tail index (e.g. extreme quantiles, small exceedance probabilities and mean excess functions) but ultimately also plays an important role in for instance the real confidence level of asymptotic confidence intervals, which are frequently used in extreme value statistics.

A somewhat more elaborate model can be obtained from expression (2) by implementing a slow variation with remainder condition ($\mathcal{R}_{b,\rho}$) on l_F, which specifies the rate of convergence to the limit in equation (3), in order to obtain bias-corrected estimates for the tail index γ. Here, we refer to Section 3.12.1 of Bingham et al. (1987).

Condition: ($\mathcal{R}_{b,\rho}$) There exists a real constant $\rho < 0$ and a rate function b satisfying $b(t) \to 0$, as $t \to \infty$, such that for all $y \geq 1$

$$\frac{\left(\frac{l_F(ty)}{l_F(t)} - 1\right)}{b(t)} \to \frac{y^\rho - 1}{\rho}, \qquad \text{as } t \to \infty. \tag{4}$$

We assume now that the slowly varying part l_F of expression (2) satisfies condition ($\mathcal{R}_{b,\rho}$) which has been widely accepted as an appropriate condition to specify the slowly varying part of a Pareto-type distribution in a semi-parametric way. Typical examples of slowly varying functions that satisfy condition ($\mathcal{R}_{b,\rho}$) are $l(y) = C + Dy^\rho(1 + o(1))$, where C and D are constants. Logarithmic functions satisfy ($\mathcal{R}_{b,0}$). In the estimation methods discussed below, the case $\rho = 0$ leads to identifiability problems (see for instance (6) below). Classical estimators can also exhibit serious bias problems in those cases.

Rewriting expression (4) in combination with (2) results in a second order refinement of limiting result (3), stating that

$$\bar{F}_{Y_t}(y) = y^{-1/\gamma}\left[1 + b(t)\frac{y^\rho - 1}{\rho} + o\big(b(t)\big)\right], \qquad \text{as } t \to \infty. \tag{5}$$

By dropping the remainder term and setting $\delta = b(t)/\rho$, this gives rise to an extension of the model obtained through expression (3), approximating the conditional distribution F_{Y_t} of relative excesses Y_t over a threshold t with

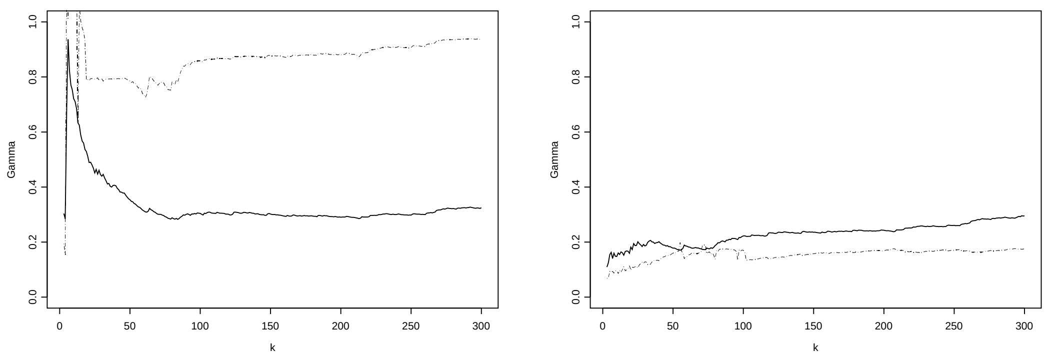
$$F_{\theta_t}(y) = \big(1 - \delta\big)\left[1 - y^{-1/\gamma}\right] + \delta\left[1 - y^{-1/\gamma+\rho}\right], \tag{6}$$

where $\theta_t = (\gamma, \delta, \rho)$ and the range of parameters is $\frac{1}{\gamma\rho} \leq \delta < 1$, $\gamma > 0$ and $\rho < 0$, in order for F_{θ_t} to be a real distribution function. Notice that the mixture model (6) is a generalized mixture model, in the sense that the weight parameter δ can also be negative and as such does not necessarily have to be located between zero and one.

Again, assuming this Pareto-mixture approximation to hold exactly as a model for the conditional distribution of the relative excesses Y_{jk} $(j = 1, \ldots, k)$ above a high threshold $t = X_{n-k,n}$ quite naturally leads to a maximum likelihood estimator for γ. This recently proposed mixture model and corresponding maximum likelihood estimator lead to potentially less biased and more stable estimates for γ than the Hill estimator. The special case, where $\rho = -1$, can be seen to be asymptotically equivalent to the well-known generalized Pareto distribution model. For more information on an asymptotically equivalent extended generalized Pareto distribution model (EGPD), we refer to Beirlant et al. (2004b).

2.3 Condroz data

In Vandewalle et al. (2004), making use of an outlier detection method based on an exponential regression model for scaled log-spacings, it was argued that the last six observations that do not immediately follow the ultimate linearity of the rest of the Pareto quantile plot in Figure 1(b), can be flagged as potential outliers under the Pareto-type model. Figure 2 shows the Hill (solid line) and bias-corrected maximum likelihood (broken line) estimates for the tail index of the Condroz data as a function of k, (a) for the entire data and (b) after rejection of the suspicious observations. On these plots, it can clearly be seen that especially the bias-corrected estimates are very sensitive to the suspicious Ca measurements. For values up to $k = 200$, also the Hill estimates drop down about 30% or more after deletion of the suspected outliers.



(a) (b)

Figure 2. (a) Hill estimates (solid line) and bias-corrected maximum likelihood estimates (broken line) for the Condroz data as a function of k and (b) same estimates after rejection of the suspicious observations.

The existence of these outliers can be explained in a number of ways, which all come down to the fact that some of the individual samples originate from regions in the community where types of soil are detected with rather peculiar characteristics. Causes for this can be found in excessive use of quicklime (calcium oxide) resulting in nuggets of calcium, the use of fertilizers by private individuals on lawns which are encoded in the databases as farmland or simply because of soil samples that are taken near the border of the community.

3 ROBUST EXTREME VALUE STATISTICS

In order to obtain more robust estimates for the tail index, the class of minimum distance estimators can be considered, which – in contrast to the class of maximum likelihood estimators – is known to have excellent robustness properties (e.g. Donoho and Liu, 1988). More specifically, in this section, we adopt the so-called partial density component estimation extension of the integrated squared error approach, as discussed for instance in Scott (2004), and apply it to the mixture approximation of equation (6) for the conditional distribution of relative excesses over a high threshold.

3.1 Partial density component estimation

The use of a minimum distance criterion based on integrated squared errors, as an alternative to maximum likelihood, was first proposed by Terrel (1990). The criterion was re-used and discussed, among others, by Hjort (1994) and Scott (1998), who demonstrated its consistency and asymptotic normality. Later, together with the maximum likelihood method, it was included in a more general framework of minimum-divergence estimators (Basu et al., 1998). It is stated that, intuitively speaking, this estimator tries to find the largest proportion of the data that matches the assumed parametric model, hence diminishing the effect of contamination.

Assuming a parametric family of distributions $F(.|\theta)$, we seek to find the parameter $\hat{\theta}$ which brings the density $f(.|\hat{\theta})$ closest to the true unknown density f, underlying the data, using an integrated squared error distance criterion:

$$\hat{\theta} = \arg\min_{\theta} \left[\int \left(f(y|\theta) - f(y) \right)^2 dy \right] \tag{7}$$

$$= \arg\min_{\theta} \left[\int f^2(y|\theta) dy - 2 \int f(y|\theta) f(y) dy + \int f^2(y) dy \right].$$

Since the middle term denotes the expected value of the model density and the minimizing value for θ does not depend on the last integral in the second line, which is constant with respect to θ, expression (7) can be rewritten as

$$\hat{\theta} = \arg\min_{\theta} \left[\int f^2(y|\theta)dy - 2E[f(Y|\theta)] \right],$$

with Y a random variable with density f.

Starting from a random sample $Y_1, \ldots, Y_k$ from Y and using the empirical mean as an unbiased estimator of $E[f(Y|\theta)]$, this gives rise to the so-called integrated squared error estimator

$$\hat{\theta}_{k,I} = \arg\min_{\theta} \left[\int f^2(y|\theta)dy - \frac{2}{k}\sum_{j=1}^{k} f(Y_j|\theta) \right]. \tag{8}$$

For many models, the first integral can be computed in closed form as a function of θ. Rewriting equation (8) as

$$\hat{\theta}_{k,I} = \arg\min_{\theta} \left[\frac{1}{k}\sum_{j=1}^{k} \left(\int f^2(y|\theta)dy - 2f(Y_j|\theta) \right) \right],$$

it is easily seen that this estimation method falls in the well-known class of M-estimators (e.g. Hampel et al., 1986) and density power divergence estimators (Basu et al., 1998). For details on this and consistency, asymptotic normality and robustness properties of the estimator, we refer to Hjort (1994), Scott (1998), Basu et al. (1998) and Juárez and Schucany (2004) in the context of parametric generalized Pareto estimation.

Perhaps more interestingly, as stated in Scott (2004), a closer examination of the arguments leading up to expression (8) reveals the fact that only f is supposed to be a real density function, whereas $f(.|\theta)$ is not. Hence, instead of using a complete density model $f(.|\theta)$ in the ISE algorithm, also an incomplete mixture model $wf(. \mid \theta)$ can be considered, i.e.

$$\hat{\theta}^{w}_{k,P} = \arg\min_{\theta,w} \left[w^2 \int f^2(y|\theta)dy - \frac{2w}{k}\sum_{j=1}^{k} f(Y_j|\theta) \right]. \tag{9}$$

This gives rise to a so-called partial density component (PDC) estimator $\hat{\theta}^{w}_{k,P} = (\hat{\theta}_{k,P}, \hat{w}_k)$, which we will be using in connection to the mixture approximation for the conditional distribution of relative excesses over a high threshold. For more information on simulation results pertaining to this new estimator, we refer to Scott (2004).

In order to somewhat better understand the usefulness of the parameter w, we differentiate the expression between brackets in equation (9) with respect to w and evaluate it in $\hat{\theta}^w_{k,P}$. Equating to zero and solving for $\hat{w}_k$ ultimately leads to

$$\hat{w}_k = \frac{1}{k}\sum_{j=1}^{k} f(Y_j|\hat{\theta}_{k,P}) \Big/ \int f^2(y|\hat{\theta}_{k,P})dy. \tag{10}$$

With F_k the empirical distribution function of the data $Y_1, \ldots, Y_k$, given by $\frac{1}{k}\sum_{j=1}^{k} \Delta_{Y_j}$, where Δ_y represents the point mass 1 in y, it is easily seen that expression (10) can be rewritten as

$$\hat{w}_k = \int f(y|\hat{\theta}_{k,P})dF_k(y) \Big/ \int f^2(y|\hat{\theta}_{k,P})dy.$$

In the following, we consider an ideal situation, in that the true density of the uncontaminated part of the data belongs to the class $F(.|\theta)$ of parametric models. Then, suppose the sample $Y_1, \ldots, Y_k$ to consist of i.i.d. observations from a contamination model with distribution

$$F^\epsilon(.) = (1-\epsilon)F(.|\theta_0) + \epsilon G(.) \tag{11}$$

where $0 \leq \epsilon < 1/2$. That is, asymptotically speaking, a proportion $(1-\epsilon)$ of the data follows distribution $F(.|\theta_0)$, within the class of parametric models, and a proportion ϵ of the data follows a contamination distribution G.

Assuming $\hat{\theta}_{k,P}$ to be a robust and consistent estimator of θ_0 and using the fact that F_k tends to F^ϵ as $k \to \infty$, it is easily seen that, heuristically speaking,

$$\hat{w}_k \approx \int f(y|\theta_0)dF^\epsilon(y) \Big/ \int f^2(y|\theta_0)dy,$$

for k large. With F^ϵ the contamination distribution as in expression (11), this results in

$$\hat{w}_k \approx (1-\epsilon) + \epsilon \int f(y|\theta_0)dG(y) \Big/ \int f^2(y|\theta_0)dy. \tag{12}$$

We emphasize here that in fact hardly any results are known yet with respect to the asymptotic behavior of the estimates (see Scott, 2004).

Still, when there is no contamination in the data ($\epsilon = 0$), i.e. the entire sample $Y_1, \ldots, Y_k$ follows a distribution $F(.|\theta_0)$, in equation (12) we see that $\hat{w}_k \approx 1$ for k large. When the data is contaminated ($\epsilon \neq 0$), the second term in expression (12) does not disappear in a natural way. Nevertheless, when G is a contamination distribution taking most of its mass outside the range of the model distribution or in a region where the density $f(.|\theta_0)$ is believed to be almost zero, the second term in equation (12) can be seen to be negligible. Practically speaking, this means that if the estimates show a robust and consistent finite sample behavior, $\hat{w}_k$ can be seen as a measure for the proportion $(1-\epsilon)$ of uncontaminated data in the sample.

Important to note however is that this somewhat heuristic interpretation concerning the $\hat{w}_k$ estimates only holds true when the proportion $(1-\epsilon)$ of good data follows the distribution $F(.|\theta_0)$ and the proportion ϵ of contamination data follows a distribution G which is taking its mass outside the range of the model distribution. In cases where the proportion of good data is only approximately $F(.|\theta_0)$ distributed and/or in cases where the proportion of contamination data follows a distribution taking most of its mass inside the range of the model distribution, further distortion in the estimates will appear. As such, in practice, care should always be taken with the interpretation of $\hat{w}_k$ as a measure for the proportion of uncontaminated data in the sample.

3.2 Tail index estimation

Following expression (6), we apply the above partial density component estimation method to the parametric family of distributions $F(.|\theta) = F_\theta(.)$ with density given by

$$f_\theta(y) = \Big(1-\delta\Big)\left[\frac{1}{\gamma}y^{-(1+1/\gamma)}\right] + \delta\left[\Big(\frac{1}{\gamma}-\rho\Big)y^{-(1+1/\gamma-\rho)}\right], \tag{13}$$

where $\theta = (\gamma, \delta, \rho)$ and $\gamma > 0$, $\rho < 0$ and $\frac{1}{\gamma\rho} \leq \delta < 1$. In Section 2.2, it was discussed how this family of distributions approximates the conditional distribution F_{Y_t} of relative excesses $Y_t = \Big(\frac{X}{t} \;\Big|\; X > t\Big)$ over a threshold t for Pareto-type distributed variables X. One can also use model (3) instead of (6); Simulation experiments have shown however that the corresponding results are severely biased and inferior with respect to mean squared error.

Consider a sample $X_1, \ldots, X_n$ of i.i.d. random variables with common distribution function F_X and denote the corresponding order statistics by $X_{1,n} \leq \ldots \leq X_{n,n}$. With $t = X_{n-k,n}$ a suitable threshold and $Y_{jk} = X_{n-j+1,n}/t$ $(j = 1, \ldots, k)$ the k corresponding relative excesses, equation (9) can be rewritten as

$$\hat{\theta}_k^w = \arg\min_{\theta,w}\left[w^2\int_1^\infty f_\theta^2(y)dy - \frac{2w}{k}\sum_{j=1}^{k} f_\theta(Y_{jk})\right]. \tag{14}$$

Here, k is easily seen to be equal to $n(1-F_n(t))$ if $t = X_{n-k,n}$, where F_n is the empirical distribution function of the original data. Since the integral $\int_1^\infty f_\theta^2(y)dy$ can be calculated easily in closed form as

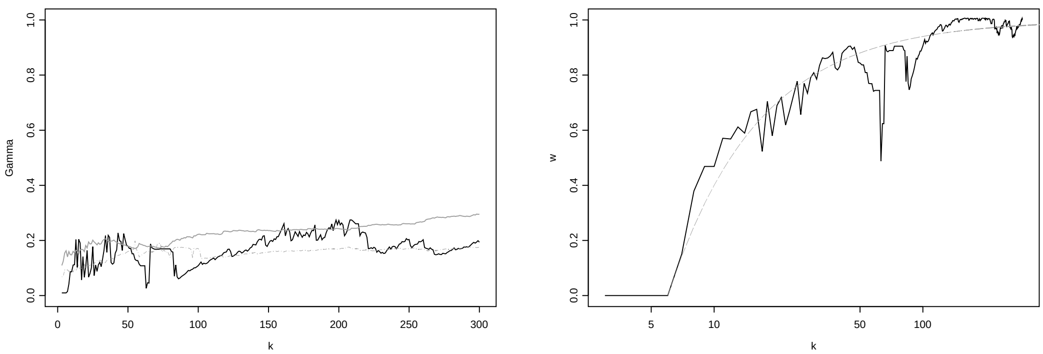
$$\varrho_1(\theta) = \left[\frac{(1+\delta)^2}{(\gamma+2)\gamma} + 2\frac{\delta(1-\delta)(1-\gamma\rho)}{(\gamma+2-\gamma\rho)\gamma} + \frac{\delta^2(1-\gamma\rho)^2}{(\gamma+2-2\gamma\rho)\gamma}\right], \tag{15}$$

this quite naturally gives rise to a partial density component estimator $\hat{\theta}_k^w = (\hat{\theta}_k, \hat{w}_k)$ for the parametric family of distributions F_θ.

3.3 Condroz data

With respect to the practical implementation of the method, an algorithm was constructed in *SPlus* to search the optimal values for the tail index γ and weight parameter w next to the second order parameters δ and ρ. The *SPlus* code is available upon request from the authors. Much in the spirit of Beirlant et al. (1999), in order to avoid instabilities in the numerical minimization routine, for instance when the optimization procedure ends at a local minimum, we introduced smoothness conditions linking estimates at subsequent values of k: given $\hat{\delta}_{k+1}$ and $\hat{\rho}_{k+1}$, we set $|\hat{\delta}_k| \leq 1.1|\hat{\delta}_{k+1}|$ and $1.1\hat{\rho}_{k+1} \leq \hat{\rho}_k \leq 0.9\hat{\rho}_{k+1}$.

Next to this, some practical bounds on the estimates of the weight parameter w could also be considered, as for instance $0 \leq \hat{w}_k \leq 1$. Nevertheless, continuing the discussion concerning $\hat{w}_k$ at the end of Section 3.1, for large values of k, the assumption that the relative excesses behave exactly as ordered data from a Pareto-mixture distribution (6), might still be over-optimistic. As such, in the following, we have considered only a lower bound $0 \leq \hat{w}_k$, which explains the possible occurrence of estimates $\hat{w}_k > 1$, for larger values of k.



(a) (b)

Figure 3. Partial density component estimates following equation (14) of (a) the tail index γ (together with the estimates from Figure 2(b) after deleting the six suspicious observations, in grey) and (b) the parameter w (log-scale for k, together with the function $(k-6)/k$, in grey), for the Condroz data, as a function of k.

Figure 3(a) shows γ estimates for the entire Condroz data, calculated with the newly proposed partial density component estimation method. The estimates are based on thresholds $X_{n-k,n}$ and are given as a function of k. Superimposed in grey are the maximum likelihood estimates of the tail index as in Figure 2(b). These were based on the data after rejection of the last six observations, which did not follow the apparent linearity of the rest of the data in the Pareto quantile plot of Figure 1(b). We see that the partial density component estimates for γ are indeed robust as they are similar to the maximum likelihood estimates based on outlier-free data.

The corresponding w estimates are shown in Figure 3(b), on a log-scale for the k-values. Also given, in grey, is the function $(k-6)/k$, representing the proportion of uncontaminated data for the relative excesses Y_{jk} from expression (14), under the assumption that the last six observations indeed correspond to contamination in the sample. Through both plots, the suggestion of the six largest observations in the data being outliers with respect to the assumption of Pareto-type distributed data is strenghtened. This is in agreement with the results as obtained in Vandewalle et al. (2004).

4 INFLUENCE FUNCTION

To study the robustness of our new estimator theoretically, we will derive its influence function. The influence function of a multivariate functional T at a distribution F describes the effect on the estimate T of an infinitesimal contamination to F at the point x and is given by

$$IF(x,T,F) = \lim_{\epsilon \to 0} \frac{T\left((1-\epsilon)F + \epsilon \Delta_x\right) - T(F)}{\epsilon}.$$

In the following, we will derive the influence function for the functional form of the newly proposed estimator in an analogous way as for the classical M-estimators (Hampel et al., 1986). General results on influence functions of the minimum density power divergence estimators can be found in Basu et al. (1998), whereas Juárez and Schucany (2004) present influence functions for parametric modelling with the generalized Pareto distribution. Here, some new results on influence functions for *partial density component* estimators are provided.

As seen before, in extreme value statistics, a popular choice for the threshold t in threshold based methods is $X_{n-k,n}$, the kth largest observation of the sample. With F_n^{-1} the empirical quantile function of the data, this threshold is easily seen to be equal to $F_n^{-1}\left(1-\frac{k}{n}\right)$. More generally speaking, to each $p \in (0,1)$ a threshold $F_n^{-1}(1-p)$ can be associated, being equal to $X_{n-k,n}$ with $k = \lfloor np \rfloor$. Then, considering the corresponding k relative excesses $Y_{jk} = X_{n-j+1,n}/X_{n-k,n}$ $(j=1,\ldots,k)$, expression (14) can be rewritten as

$$\hat{\theta}_p^w = \arg\min_{\theta,w} \left[\frac{1}{k} \sum_{j=1}^{k} \varrho\left(X_{n-j+1,n}, F_n^{-1}(1-p), \theta, w\right) \right],$$

with ϱ given by $\varrho(x,t,\theta,w) = w^2\varrho_1(\theta) - w\varrho_2(x,t,\theta)$. Here, $\varrho_2(x,t,\theta) = 2f_\theta\left(\frac{x}{t}\right)$ and ϱ_1 is defined as in equation (15).

Now, as k/n does not depend on θ or w, it is easily seen that $\hat{\theta}_p^w$ can also be written as

$$\hat{\theta}_p^w = \arg\min_{\theta,w} \left[\frac{1}{n} \sum_{j=1}^{k} \varrho \left(X_{n-j+1,n}, F_n^{-1}(1-p), \theta, w \right) \right].$$

Using the empirical distribution function $F_n = \frac{1}{n} \sum_{j=1}^{n} \Delta_{X_{n-j+1,n}}$ of the data, this allows us to define the estimator as the value $\hat{\theta}_p^w = (\hat{\theta}_p, \hat{w}_p) = (\hat{\gamma}_p, \hat{\delta}_p, \hat{\rho}_p, \hat{w}_p)$ which minimizes

$$\int_{F_n^{-1}(1-p)}^{\infty} \varrho \left(x, F_n^{-1}(1-p), \theta^w \right) dF_n(x)$$

with respect to $\theta^w = (\theta, w)$. Or, with $\psi(x, t, \theta^w) = \partial \varrho(x, t, \theta^w)/\partial \theta^w$ representing the vector function of partial derivatives of ϱ with respect to the parameters of θ^w, as the solution $\hat{\theta}_p^w$ of the vector equation

$$\int_{F_n^{-1}(1-p)}^{\infty} \psi \left(x, F_n^{-1}(1-p), \hat{\theta}_p^w \right) dF_n(x) = 0.$$

As such, the estimator $\hat{\theta}_p^w$ can be written as $T_p(F_n)$, where T_p is a functional given by

$$\int_{F^{-1}(1-p)}^{\infty} \psi \left(x, F^{-1}(1-p), T_p(F) \right) dF(x) = 0. \tag{16}$$

We continue by rewriting equation (16) for the contaminated distribution as given in the definition of the influence function, i.e.

$$\begin{aligned} (1-\epsilon) \int_{F_\epsilon^{-1}(1-p)}^{\infty} \psi \left(z, F_\epsilon^{-1}(1-p), T_p(F_\epsilon) \right) dF(z) \\ + \ \epsilon \ \psi \left(x, F_\epsilon^{-1}(1-p), T_p(F_\epsilon) \right) I_{\left\{ x \geq F_\epsilon^{-1}(1-p) \right\}} = 0 \end{aligned} \tag{17}$$

with $F_\epsilon = (1-\epsilon)F + \epsilon \Delta_x$. To compute $IF(x, T_p, F) = \partial T_p(F_\epsilon)/\partial \epsilon |_{(\epsilon=0)}$ we differentiate expression (17) with respect to ϵ and take the limit for ϵ going to 0. The second term in expression (17) has partial derivative

$$\begin{aligned} \frac{\partial}{\partial \epsilon} \left[\epsilon \psi \left(x, F_\epsilon^{-1}(1-p), T_p(F_\epsilon) \right) I_{\{x \geq F^{-1}(1-p)\}} \right] \Big|_{(\epsilon=0)} \\ = \ \psi \left(x, F^{-1}(1-p), T_p(F) \right) I_{\{x \geq F^{-1}(1-p)\}}, \end{aligned} \tag{18}$$

whereas for the first term, we obtain

$$\frac{\partial}{\partial\epsilon}\Big[(1-\epsilon)\int_{F_\epsilon^{-1}(1-p)}^{\infty}\psi\left(z,F_\epsilon^{-1}(1-p),T_p(F_\epsilon)\right)dF(z)\Big]\Big|_{(\epsilon=0)} \tag{19}$$

$$= \frac{\partial}{\partial\epsilon}\Big[\int_{F_\epsilon^{-1}(1-p)}^{\infty}\psi\left(z,F_\epsilon^{-1}(1-p),T_p(F_\epsilon)\right)dF(z)\Big]\Big|_{(\epsilon=0)}$$

$$- \int_{F^{-1}(1-p)}^{\infty}\psi\left(z,F^{-1}(1-p),T_p(F)\right)dF(z).$$

With the definition of T_p as in equation (16), the last term in equation (19) disappears. The first term however, using Leibnitz's integral rule, transforms to

$$\frac{\partial}{\partial\epsilon}\Big[\int_{F_\epsilon^{-1}(1-p)}^{\infty}\psi\left(z,F_\epsilon^{-1}(1-p),T_p(F_\epsilon)\right)dF(z)\Big]\Big|_{(\epsilon=0)} \tag{20}$$

$$= \int_{F^{-1}(1-p)}^{\infty}\frac{\partial}{\partial\epsilon}\Big[\psi\left(z,F_\epsilon^{-1}(1-p),T_p(F_\epsilon)\right)\Big]\Big|_{(\epsilon=0)}dF(z)$$

$$- \psi\left(F^{-1}(1-p),F^{-1}(1-p),T_p(F)\right)f(F^{-1}(1-p))$$

$$\times \frac{\partial F_\epsilon^{-1}(1-p)}{\partial\epsilon}\Big|_{(\epsilon=0)}.$$

It is not so difficult to see that ultimately, further evaluating the first term in equation (20), one obtains

$$\frac{\partial}{\partial\epsilon}\Big[\int_{F_\epsilon^{-1}(1-p)}^{\infty}\psi\left(z,F_\epsilon^{-1}(1-p),T_p(F_\epsilon)\right)dF(z)\Big]\Big|_{(\epsilon=0)} \tag{21}$$

$$= M(T_p(F),F)\frac{\partial T_P(F_\epsilon)}{\partial\epsilon}\Big|_{(\epsilon=0)} + \kappa(T_p(F),F)\frac{\partial F_\epsilon^{-1}(1-p)}{\partial\epsilon}\Big|_{(\epsilon=0)}$$

with M the 4×4 matrix and κ the vector given by

$$M(T_p(F),F) = \int_{F^{-1}(1-p)}^{\infty}\frac{\partial}{\partial\theta}\Big[\psi\left(z,F^{-1}(1-p),\theta\right)\Big]\Big|_{(\theta=T_p(F))}dF(z) \tag{22}$$

$$\kappa(T_p(F),F) = \int_{F^{-1}(1-p)}^{\infty}\frac{\partial}{\partial t}\Big[\psi\left(z,t,T_p(F)\right)\Big]\Big|_{(t=F^{-1}(1-p))}dF(z)$$

$$- \psi\left(F^{-1}(1-p),F^{-1}(1-p),T_p(F)\right)f(F^{-1}(1-p)).$$

Combining equations (18), (19), (21) and (22) and using the fact that

$$\left.\frac{\partial F_\epsilon^{-1}(1-p)}{\partial \epsilon}\right|_{(\epsilon=0)} = IF(x, F^{-1}(1-p), F) = \frac{I_{\{x\geq F^{-1}(1-p)\}} - p}{f(F^{-1}(1-p))}$$

for any absolute continuous distribution function F with density function f and $f(F^{-1}(1-p)) \neq 0$ (see Huber, 1981) eventually leads to the system of influence functions given by

$$IF(x, T_p, F) = -M(T_p(F), F)^{-1}\Psi\left(x, T_p(F), F\right), \tag{23}$$

with Ψ defined as

$$\begin{aligned}\Psi\left(x, T_p(F), F\right) &= \psi\left(x, F^{-1}(1-p), T_p(F)\right) I_{\{x\geq F^{-1}(1-p)\}} \\ &+ \ \kappa(T_p(F), F)\frac{I_{\{x\geq F^{-1}(1-p)\}} - p}{f(F^{-1}(1-p))}.\end{aligned}$$

Examples

Concerning the parameter w, notice how in expression (16), for the partial derivative with respect to w, we obtain

$$\int_{F^{-1}(1-p)}^{\infty} \left[2w\varrho_1(\theta_p) - \varrho_2\left(x, F^{-1}(1-p), \theta_p\right)\right] dF(x) = 0,$$

which, after solving for w, results in

$$w = \frac{1}{2p\rho_1(\theta_p)}\int_{F^{-1}(1-p)}^{\infty} \varrho_2\left(x, F^{-1}(1-p), \theta_p\right) dF(x).$$

From this, together with the parameters γ, δ and ρ from the asymptotic expansions (4) or (5) for a Pareto-type distribution F, it is possible to plot the influence function $IF(x, T_p, F)$ as given in equation (23).

As a first example, we consider the Fréchet distribution with parameter $\alpha > 0$, given by

$$\bar{F}(x) = 1 - \exp(-x^{-\alpha}),$$

for $x > 0$. Using a Taylor expansion of $\exp(-x^{-\alpha})$, it is easily seen that $\bar{F}(x) = x^{-1/\gamma} l_F(x)$ with $\gamma = 1/\alpha$ and

$$l_F(x) = 1 - \frac{x^{-\alpha}}{2} + o\left(x^{-\alpha}\right), \qquad \text{as } x \to \infty.$$

As such, for the slow variation with remainder condition $(\mathcal{R}_{b,\rho})$ on l_F, we obtain $b(t) = \frac{\alpha}{2}t^{-\alpha}$ and $\rho = -\alpha$. Now, taking $\delta = b\left(F^{-1}(1-p)\right)/\rho$, the influence functions of T_p, as given in equation (23), can be calculated for each choice of p and α.

In Figure 4, the influence functions of T_p are shown for the Fréchet distribution with parameter $\alpha = 2$ and threshold at the upper p-quantile of the distribution, with $p = 0.25$. The influence functions are seen to be continuous, except in the threshold. For contamination to the left of the threshold, influence is negligible in all cases. The influence for the model parameters γ, δ and ρ also becomes negligible for outliers far to the right of the threshold. For all parameters, contamination close to the right side of the threshold results in the largest influence, as well as for parameter w in case of far outliers to the right. In any case, the influence functions are bounded, making the functional T_p B-robust at F (Hampel et al., 1986).

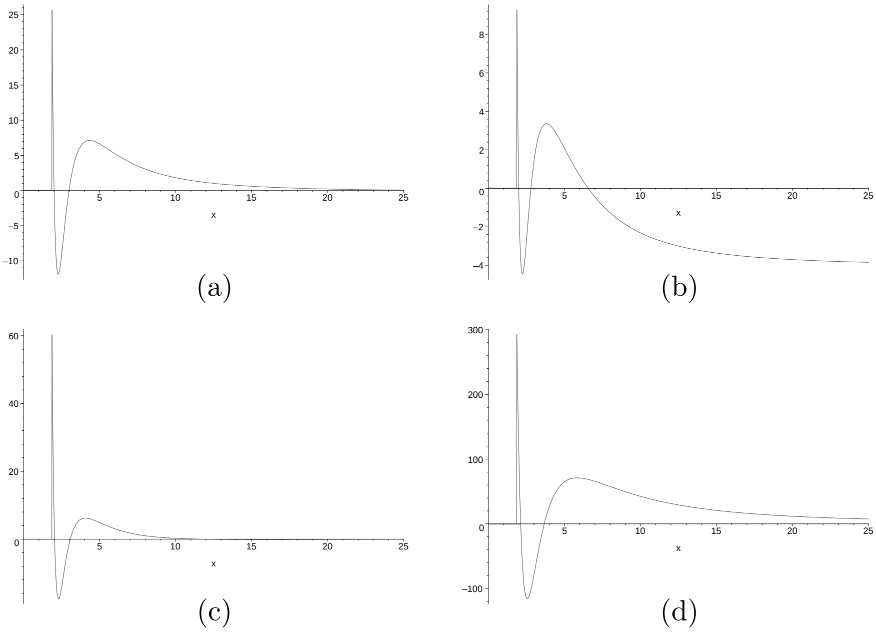


Figure 4. Influence functions of the functional T_p for (a) parameter γ, (b) parameter w, (c) parameter δ and (d) parameter ρ, based on the Fréchet distribution with α=2 and p=0.25.

A second example is the Burr distribution with parameters $\beta > 0$, $\tau > 0$ and $\lambda > 0$, given by

$$\bar{F}(x) = \left(\frac{\beta}{\beta + x^\tau} \right)^\lambda ,$$

for $x > 0$. Here, it's not so difficult to see that $\bar{F}(x) = x^{-1/\gamma} l_F(x)$ with $\gamma = 1/(\tau\lambda)$ and

$$l_F(x) = \beta^\lambda \left(\beta x^{-\tau} + 1 \right)^{-\lambda} ,$$

such that we obtain $(\mathcal{R}_{b,\rho})$ on l_F with $b(t) = \beta\tau\lambda t^{-\tau}$ and $\rho = -\tau$. Again, after taking $\delta = b\left(F^{-1}(1-p)\right)/\rho$, this allows for the influence functions of T_p to be calculated for each choice of p and β, τ and λ. Figure 5 shows the influence functions of T_p for the Burr distribution with parameters $\beta = 1$, $\tau = 1/2$

and $\lambda = 2$, and threshold at the upper p-quantile of the distribution, with $p = 0.10$. Same conclusions hold as for the influence functions of Figure 4.

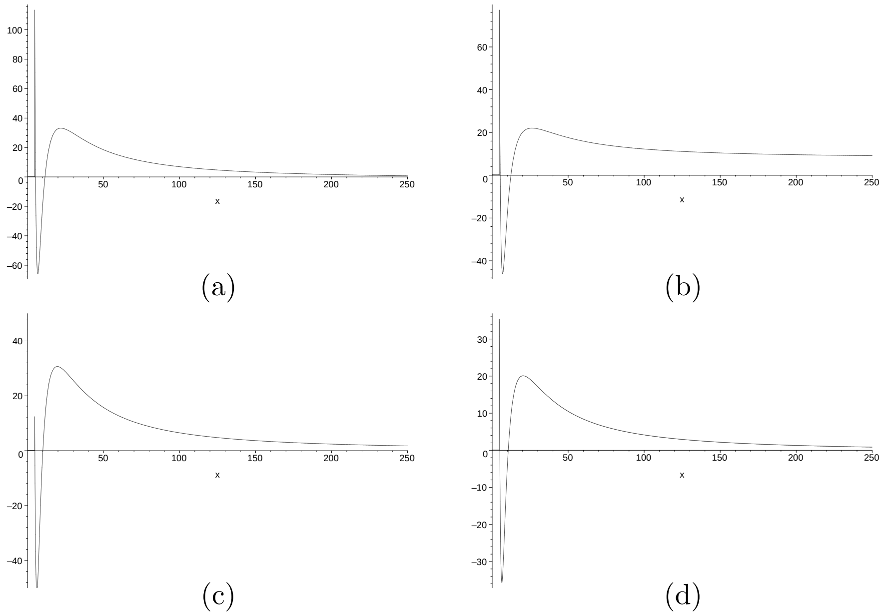


Figure 5. Influence functions of the functional T_p for (a) parameter γ, (b) parameter w, (c) parameter δ and (d) parameter ρ, based on the Burr distribution with β=1, τ=0.5, λ=2 and p=0.10.

5 SIMULATION STUDY

In order to get a better understanding of the practical usefulness of the proposed robust statistical method, its finite sample behavior is investigated, both at contaminated as well as uncontaminated data. For the tail index γ, a comparison is made between the well-known Hill (1975) estimator and the newly proposed partial density component estimator.

5.1 Behavior at uncontaminated data

We consider 25 simulated samples without contamination, each containing $n = 200$ observations, from the two Pareto-type distributions F as already discussed before:

- The Fréchet distribution given by $\bar{F}(x) = 1 - \exp(-x^{-\alpha})$, such that $\rho = -\alpha$ and $\gamma = 1/\alpha$.

- The Burr distribution given by $\bar{F}(x) = \left(\frac{\beta}{\beta+x^\tau}\right)^\lambda$, such that $\rho = -\tau$ and $\gamma = 1/(\lambda\tau)$.

In the simulations, we have chosen $\alpha = 2$ for the Fréchet distribution and $\beta = 1$, $\tau = 2$ and $\lambda = 1$ for the Burr distribution. In the top panels of Figure 6, the means of the Hill (1975) estimator (dash-dotted line) and bias corrected estimator (dashed line) based on mixture model (6) are superimposed, in grey, onto the means of the partial density component estimator, as a function of k. The horizontal line indicates the true value of γ, in both cases equal to 1/2. The middle panels of Figure 6 show the corresponding empirical mean squared errors, also as a function of k. Finally, in the bottom panels of Figure 6 also the mean w estimates, as obtained through the partial density component estimator, are shown as a function of k.

Although most robust estimators are known to be less efficient at the uncontaminated model than maximum likelihood estimators, we notice that the newly proposed partial density component estimator still performs relatively well in case of uncontaminated data. The γ estimates still seem to be fairly stable around the true value of γ and even with respect to mean squared error, the newly proposed estimator does not seem to lose too much accuracy. Only at smaller k-values, the partial density component estimators underestimate γ and w, and the mean squared errors of the estimates increase substantially.

5.2 Behavior at contaminated data

With respect to the finite sample behavior of the estimators at contaminated data, we consider the same distributions as above, but with different amounts of contamination. Here, as contamination distribution G, we have chosen the Pareto distribution, given by $G(x) = 1 - \left(\frac{x}{x_c}\right)^{-\alpha}$, for $x \geq x_c$, with $\alpha = 1/2$. The true distribution of the data is then taken to be

$$F_X^\epsilon = (1-\epsilon)F + \epsilon G$$

with F the Fréchet distribution or Burr distribution and ϵ the amount of contamination. Concerning the location parameter x_c two different values are considered for each combination of distribution function F and amount of contamination ϵ. Either x_c is taken to correspond to 1.2 times the largest value in the uncontaminated part of the 150 considered data sets or as 2 times this value, as such representing contamination closer or further away from the uncontaminated part of the observations. For the amount of contamination itself, choice is made between 0.01, 0.025, 0.05 and 0.1 for the value of ϵ. Table 5.2 shows the different settings used in our simulations.

Table 5.2.

	Fréchet	Burr	
Fig.	x_c	x_c	ϵ
7	1.2	2.0	0.01
	2.0	1.2	0.025
	1.2	2.0	0.05
8	2.0	1.2	0.1

In Figures 7 and 8, the results of the partial density component estimator are given as follows: in the left panel the results for the Fréchet distribution are given, whereas the right panel corresponds to the Burr distribution. In the top panels, the means of the γ estimates, obtained with the partial density component estimator, are given as a function of k. Superimposed in grey are the corresponding Hill (1975) estimates (dash-dotted line). Again, the horizontal line indicates the true value of γ, corresponding to the uncontaminated part of the data. The middle panels of Figures 7 and 8 show the corresponding empirical mean squared errors, also as a function of k. Finally, in the bottom panels, the mean w estimates of the partial density component estimator are given, on a log-scale for k. Also given here is the function $(k-n_c)/k$, representing the proportion of uncontaminated data for the relative excesses Y_{jk} from expression (14), under the assumption that the last $n_c = 200\epsilon$ observations indeed correspond to the contamination in the sample.

The γ estimates seem to be fairly stable for intermediate values of k, making the influence of the choice of k less troublesome than e.g. for the Hill (1975) estimator in case of uncontaminated data. In comparison to the maximum likelihood estimators, we notice that, although there is a slight tendency to underestimation, the newly proposed partial density component estimator performs remarkably better. With respect to the w estimates, we see that for the cases with larger amounts of contamination, the estimates seem to follow the expected values $(k - n_c)/k$ relatively well for the entire range of k values. For smaller amounts of contamination and smaller values of k however, the w estimates seem to be less capable of discriminating between the good observations and the introduced outliers. With respect to the remaining settings that can be found in Table 5.2, i.e. for $\epsilon = 0.025$ and $\epsilon = 0.05$, results have not been represented graphically. Nevertheless, also in these cases, the behavior of the respective estimators shows to be consistent with the above findings.

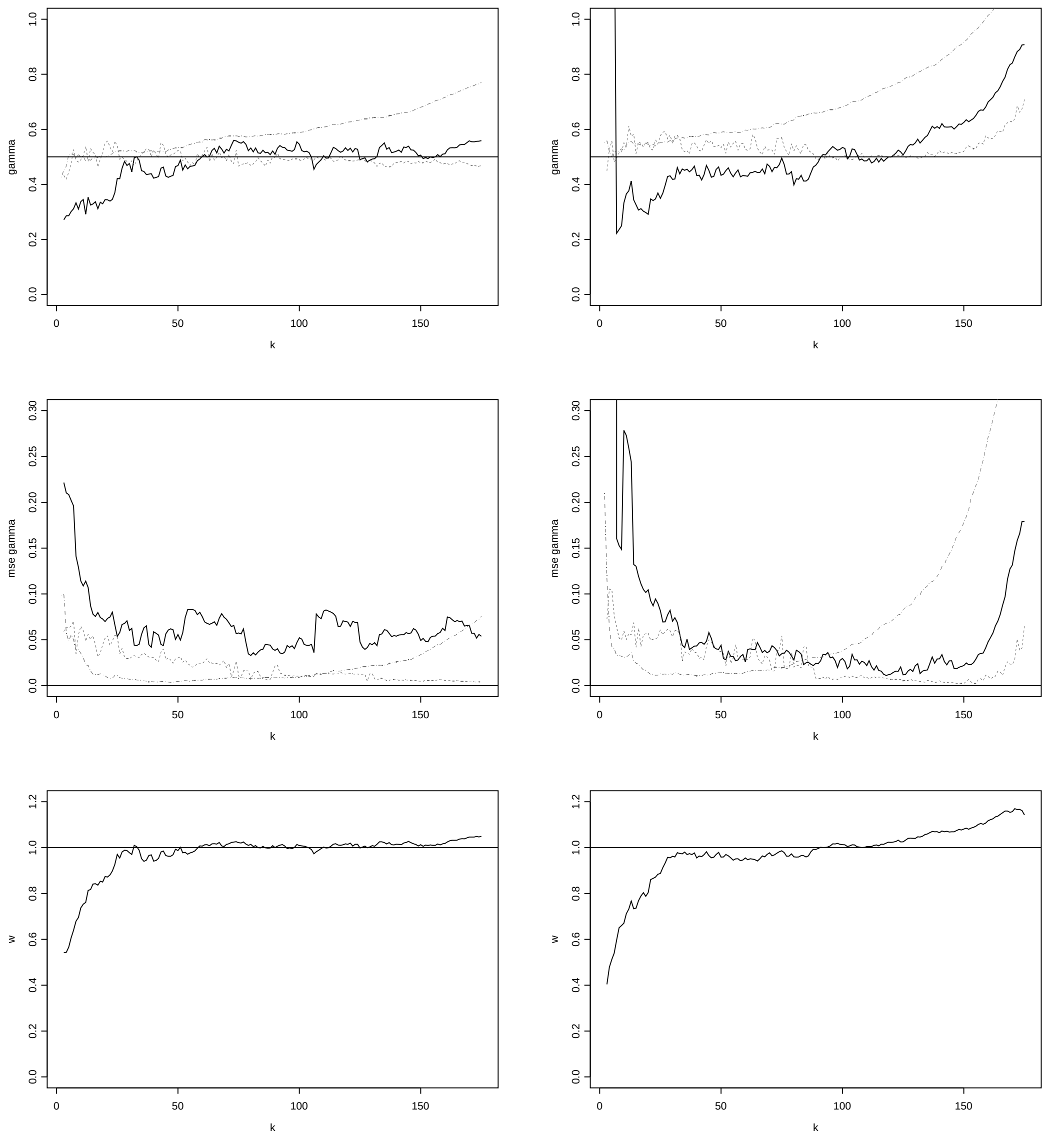


Figure 6. Mean γ estimates (top panels), corresponding empirical mean squared errors (middle panels) and w estimates (bottom panels) for the Fréchet distribution (left panels) with $\alpha = 2$ and Burr distribution (right panels) with β=1, τ=2 and λ=1 (no contamination). Superimposed are the results for the Hill estimator (dash-dotted line) and bias corrected estimator (6) (dashed lines).

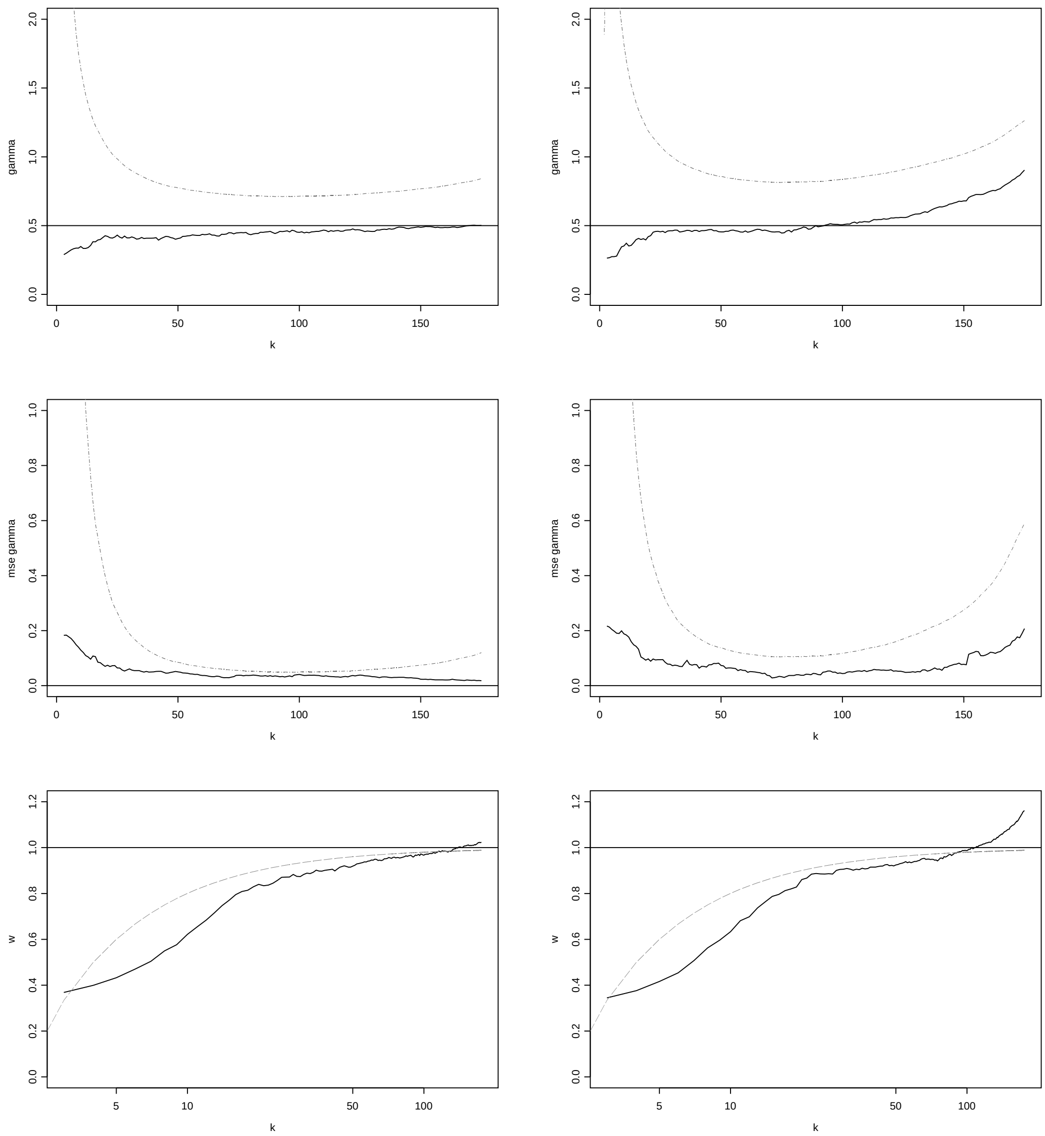


Figure 7. Mean γ estimates (top panels), corresponding empirical mean squared errors (middle panels) and w estimates (bottom panels) for the Fréchet distribution (left panels) and Burr distribution (right panels) with ϵ=0.01 as in Table 5.2. Given as a reference are the corresponding Hill (1975) estimates (top and middle panels, dash-dotted line) and the expected values of w (bottom panels, dashed line).

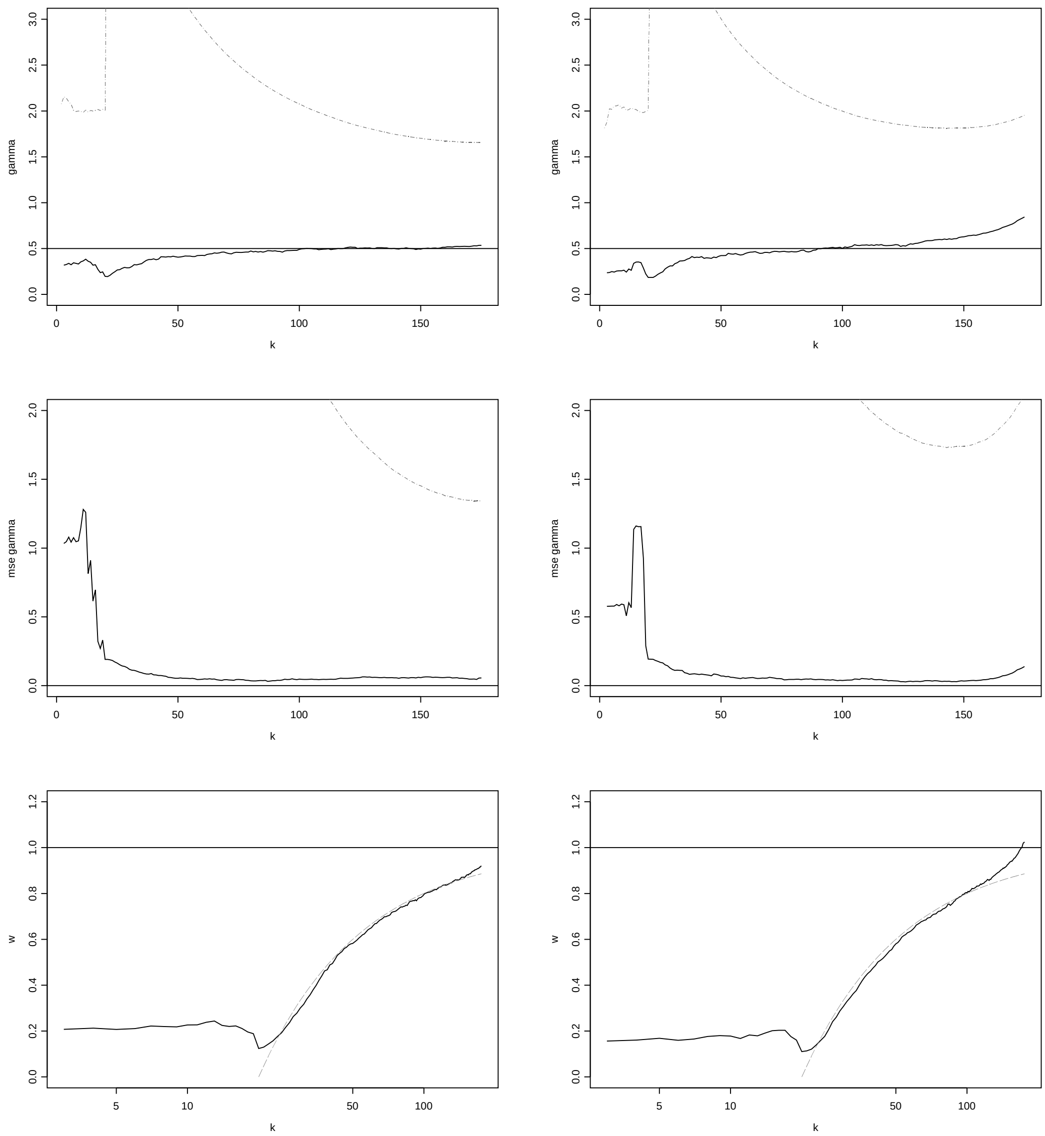


Figure 8. Mean γ estimates (top panels), corresponding empirical mean squared errors (middle panels) and w estimates (bottom panels) for the Fréchet distribution (left panels) and Burr distribution (right panels) with ϵ=0.1 as in Table 5.2. Given as a reference are the corresponding Hill (1975) estimates (top and middle panels, dash-dotted line) and the expected values of w (bottom panels, dashed line).

6 CONCLUSIONS

We have introduced a new robust estimator for the tail index of Pareto type distributions, based on a robust integrated squared error approach applied to a mixture model for relative excesses over a large threshold. Special attention was paid to the influence function of the estimator and the role of the weight parameter. When applied to Ca measurements from the Condroz region in Belgium, we could easily identify the outliers that were also seen on a Pareto quantile plot, and we obtained much lower estimates of the tail index.

Finally, in a small simulation study, the new estimator was compared to the classical ones and the reliability of the estimates was examined. From the empirical results, we consider the newly proposed robust method a useful data analytical tool that can help the practitioner obtain robust estimates for the tail index and/or check for potential outliers in case of Pareto-type distributed data.

References

Basu, A., Harris, I.R., Hjort, N.L., Jones, M.C. 1998. Robust and efficient estimation by minimising a density power divergence. Biometrika 85, 549–560.

Beirlant, J., Dierckx, G., Goegebeur, Y., Matthys, G., 1999. Tail index estimation and an exponential regression model. Extremes 2 (2), 177–200.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2004a. Statistics of Extremes - Theory and Applications. Wiley.

Beirlant, J., Joossens, E., Segers, J., 2004b. Discussion of "Generalized Pareto Fit to the Society of Actuaries' Large Claims Database" by A. Cebrian, M. Denuit and P. Lambert. North American Actuarial Journal 8, 108–111.

Beirlant, J., Vynckier, P., Teugels, J., 1996. Tail index estimation, pareto quantile plots, and regression diagnostics. Journal of the American Statistical Association 91, 1659–1667.

Bingham, N., Goldie, C., Teugels, J., 1987. Regular Variation. Cambridge University Press.

Brazauskas, V., Serfling, R., 2000a. Robust and efficient estimation of the tail index of a single-parameter Pareto distribution. North American Actuarial Journal 4, 12–27.

Brazauskas, V., Serfling, R., 2000b. Robust estimation of tail parameters for two-parameter and exponential models via generalized quantile statistics. Extremes 3, 231–249.

Donoho, D.L., Liu, R.C., 1988. The 'automatic' robustness of minimum distance functionals. The Annals of Statistics 16, 552–586.

Dupuis, D., Field, C., 1998. Robust estimation of extremes. Canadian Journal of Statistics 26 (2), 119–215.

Fisher, R., Tippett, L., 1928. On the estimation of the frequency distribution of the largest or smallest member of a sample. Proceedings of the Cambridge Philosophical Society 24, 180–190.

Goegebeur, Y., Planchon, V., Beirlant, J., Oger, R., 2005. Quality assessment of pedochemical data using extreme value methodology. Journal of Applied Science 5 (6), 1092–1102.

Hampel, F., Ronchetti, E., Rousseeuw, P., Stahel, W., 1986. Robust Statistics: The Approach Based on Influence Functions. Wiley.

Hill, B., 1975. A simple general approach to inference about the tail of a distribution. The Annals of Statistics 3, 1163–1174.

Hjort, N.L., 1994. Minimum L_2 and robust Kullback-Leibler estimation. Proceedings of the 12th Prague Conference on Information Theory, edited by P. Lachout and J.A. Víšek, Prague Academy of Sciences of the Czech Republic, 102–105.

Huber, P., 1981. Robust Statistics. Wiley.

Juárez, S., Schucany, W., 2004. Robust and efficient estimation for the generalized Pareto distribution. Extremes 7 (3), 237–251.

Kratz, M., Resnick, S., 1996. The qq-estimator of the index of regular variation. Communications in Statistics: Stochastic Models 12, 699–724.

Laroche, J., Oger, R., 1999. Base de données sols. Première synthèse. Faculté universitaire des Sciences agronomique de Gembloux. Unité de Géopédologie. Requasud asbl. Gembloux.

Peng, L., Welsh, A., 2001. Robust estimation of the generalized Pareto distribution. Extremes 4, 53–65.

Scott, D.W., 1998. Parametric modeling by minimum L_2 error. TR 98-3, Rice University, Houston.

Scott, D.W., 2004. Partial mixture estimation and outlier detection in data and regression. Theory and Applications of Recent Robust Methods, edited by M. Hubert, G. Pison, A. Struyf and S. Van Aelst, 297–306 (Birkhauser, Basel).

Smith, R., 1987. Estimating tails of probability distributions. Annals of Statistics 15, 1174–1207.

Terrel, G., 1990. Linear density estimates. Proceedings of the Statistical Computing Section, American Statistical Association, 297–302.

Vandewalle, B., Beirlant, J., Hubert, M., 2004. A robust estimator of the tail index based on an exponential regression model. Theory and Applications of Recent Robust Methods, edited by M. Hubert, G. Pison, A. Struyf and S. Van Aelst, 367–376 (Birkhauser, Basel).