
Robust Estimators in Generalized Pareto Models

Peter Ruckdeschel · Nataliya Horbenko

Received: date / Accepted: date

Abstract We study global and local robustness properties of several estimators for shape and scale in a generalized Pareto model. The estimators considered in this paper cover maximum likelihood estimators, skipped maximum likelihood estimators, Cramér-von-Mises Minimum Distance estimators, and, as a special case of quantile-based estimators, Pickands Estimator.

We further consider an estimator matching the population median and an asymmetric, robust estimator of scale (kMAD) to the empirical ones (kMedMAD), which may be tuned to an expected FSBP of 34%.

These estimators are compared to one-step estimators distinguished as optimal in the shrinking neighborhood setting, i.e.; the most bias-robust estimator minimizing the maximal (asymptotic) bias and the estimator minimizing the maximal (asymptotic) MSE. For each of these estimators, we determine the finite sample breakdown point, the influence function, as well as statistical accuracy measured by asymptotic bias, variance, and mean squared error—all evaluated uniformly on shrinking convex contamination neighborhoods. Finally, we check these asymptotic theoretical findings against finite sample behavior by an extensive simulation study.

Keywords global robustness · local robustness · finite sample breakdown point · generalized Pareto distribution

Mathematics Subject Classification (2000) MSC 62F35

This work was supported by a DAAD scholarship for N.H.

P. Ruckdeschel · N. Horbenko
Fraunhofer ITWM, Department of Financial Mathematics,
Fraunhofer-Platz 1, D-67663 Kaiserslautern
and Dept. of Mathematics, University of Kaiserslautern,
P.O.Box 3049, D-67653 Kaiserslautern
E-mail: peter.ruckdeschel@itwm.fraunhofer.de
nataliya.horbenko@itwm.fraunhofer.de

1 Introduction

The topic of this paper is robust parameter estimation in generalized Pareto distributions (GPDs). These arise naturally in many situations where one is interested in the behavior of extreme events as motivated by the Pickands-Balkema-de Haan extreme value theorem (PBHT), cf. Balkema and de Haan (1974), Pickands (1975). The application we have in mind is the calculation of the regulatory capital as required by Basel II (2006) for a bank to cover operational risk. In quantifying this risk, usually the tail behavior of the underlying distribution is crucial. Estimating these population quantiles by their empirical counterparts apparently is drastically prone to outliers. This is where extreme value theory enters, suggesting to estimate these quantiles parametrically using, e.g., GPDs, see e.g. Neslehova et al. (2006). This per se is no remedy, however. Maximum Likelihood Estimators (MLEs), optimal in this parametric context, still attribute unbounded influence to some exposed observations. Robust Statistics in contrast offers procedures bounding the influence of single observations, so provides reliable inference in the presence of moderate deviations from the underlying model assumptions, respectively the mechanisms underlying the PBHT. Admittedly, this comes at the price of some efficiency loss in the ideal model.

Literature Estimating the three-parameter GPD has been a challenging problem for statisticians for long, with many proposed approaches. The MLE for the GPD is very popular for practitioners, and has been studied in detail by Smith (1987). To study the instability of this procedure, Cope et al. (2009) consider skipping some extremal data peaks, thereby reducing the influence of extreme values. Grossly speaking this amounts to using a Skipped Maximum Likelihood Estimators (SMLE). Close to this is the weighted likelihood method proposed in Dupuis and Morgenthaler (2002). Following the general lines to obtain optimally-robust estimators, Dupuis (1998) and Dupuis and Field (1998) recommend an Optimal Bias-Robust Estimator (OBRE): to a given bound on the bias in the neighborhood, its influence function minimizes the trace of the variance (Hampel et al., 1986, 2.4 Thm. 1). Generalizing He and Fung (1997), Peng and Welsch (2001) propose a method of median estimator claimed to be very robust, which is based on solving the implicit equations matching the population medians of the coordinates of the scores function to the data. A special case of the Elementary Percentile Method (EPM) introduced by Castillo and Hadi (1997) may be seen in Pickands estimator (PE), Pickands (1975), striking out for its closed form representation. Brazauskas and Serfling (2000) use a different parametrization of the GPD, i.e.; instead of observations $X_i \stackrel{\text{i.i.d.}}{\sim} \text{GPD}(\beta, \xi)$ in our notation, consider observations $Y_i = X_i + \beta/\xi$ and parametrize their model by $\alpha = \xi^{-1}$ and $\sigma = \beta/\xi$. In their setting, $\mathcal{L}(\log(Y_i)) = \mathcal{L}(\log(\beta/\xi) + E/\xi)$, $E \sim \text{Exp}(1)$, so they can transform the problem to a location-scale problem for the exponential distribution. In our setting though, their procedures are not directly applicable, as β/ξ is unknown. Other approaches cover the Method of Moments and the Method of Probability Weighted Moments (Hosking and Wallis, 1987) and Minimum Density Power Divergence (distance) Estimator (Juárez, 2003; Juárez and Schucany, 2004). We do not study these estimators here, though.

Estimators considered in this paper (for actual definitions see section 2):

- the Maximum Likelihood Estimator (MLE)

- the Skipped Maximum Likelihood Estimator (SMLE)
- the Cramér-von-Mises Minimum Distance estimator (MDE)
- Pickands Estimator (PE)
- the Method-of-Median estimator (MMed)
- an estimator based on median and kMAD (kMedMAD)
- the most bias-robust estimator minimizing the maximal bias (MBRE)
- the estimator minimizing the maximal MSE (OMSE)

MLE, MBRE, and OMSE are optimal in certain settings, so serve as benchmarks. PE, MMed, and kMedMAD are candidates for (robust) initialization estimators, and SMLE, MDE are competitors in our application.

We compare these estimators as to standard local and global robustness properties as well as by efficiencies in the ideal model and on suitable neighborhoods.

Remark 1.1 This paper is a part of the PhD thesis of the second author; a preliminary version of it is Ruckdeschel and Horbenko (2010a), abbreviated henceforth R. resp. H. It contains additional tables and figures and covers, in addition, moment-based estimators, kMedMAD for $k = 1$, and variants of Pickands estimator tuned for optimal FSBP (in the class of PE-type estimators) and better variance. These estimators though have not been convincing and hence are left out here.

Structure of the paper In Sections 1.1 and 1.2, we outline the generalized Pareto distribution, define contamination neighborhoods, and recall global (finite sample breakdown point) and local (influence function) robustness criteria for estimators, together with accuracy measures such as asymptotic bias, variance, and mean squared error (MSE). Section 2 gathers the robustness properties of the above-mentioned estimators: We analytically calculate the influence functions, breakdown points and asymptotic accuracy measures for MLE, SMLE, PE, kMedMAD, and MDE, and, numerically, for MMed, MBRE, and OMSE estimators.

Our contribution is the kMedMAD estimator which improves the “initialization-free” estimators known so far considerably. Also, in the GPD context, MBRE and OMSE have not yet been compared to the cited estimators as to their asymptotic variances, and maximal MSEs. Another important contribution of this paper is a synopsis Section 3 where in tables and graphics we summarize our findings at a representative reference parameter setting; (see also Figure 3. and Table 3, though). A simulation study in Section 4 checks for the validity of the theoretical concepts, so far all based on asymptotics, i.e.; for sample size n tending to infinity. In contrast to other approaches, for realistic comparisons, we allow for estimator-specific contamination such that each estimator has to prove its usefulness in its individual worst contamination situation. This is particularly important for estimators with redescending influence function, where drastically large observations will not be the worst situation to produce bias. The conclusions from our findings are summarized in Section 5.

1.1 Model Setting

Generalized Pareto Distribution The three-parameter generalized Pareto distribution (GPD) has c.d.f. and density

$$F_{\theta}(x) = 1 - \left(1 + \xi \frac{x - \mu}{\beta}\right)^{-\frac{1}{\xi}}, \quad f_{\theta}(x) = \frac{1}{\beta} \left(1 + \xi \frac{x - \mu}{\beta}\right)^{-\frac{1}{\xi} - 1} \quad (1.1)$$

where $x \geq \mu$ for $\xi \geq 0$, and $\mu < x \leq \mu - \frac{\beta}{\xi}$ if $\xi < 0$. It is parametrized by $\theta = (\xi, \beta, \mu)^\tau$, for location μ , scale $\beta > 0$ and shape ξ . Special cases of GPDs are the uniform ($\xi = -1$), the exponential ($\xi = 0, \mu = 0$), and Pareto ($\xi > 0, \beta = 1$) distributions. We limit ourselves to the case shape $\xi > 0$ and known location $\mu = 0$ here.

GPD is a good candidate for modeling the distributional tails from the threshold point μ on as motivated by the PBHT.

For all graphics and both numerical and simulational evaluations we use the reference parameter values $\beta = 1$ and $\xi = 0.7$.

For known μ , the model is smooth, i.e. L_2 -differentiable, as the density f_θ is differentiable in θ and the corresponding Fisher information is finite and continuous in θ (Witting, 1985, Satz 1.194), with L_2 -derivative

$$\Lambda_\theta(z) = \left(\frac{1}{\xi^2} \log(1 + \xi z) - \frac{\xi+1}{\xi} \frac{z}{1+\xi z}, -\frac{1}{\beta} + \frac{\xi+1}{\beta} \frac{z}{1+\xi z} \right)^\tau, \quad z = \frac{x-\mu}{\beta} \quad (1.2)$$

For integrations it turns out useful to introduce

$$v^{-\xi} = 1 + \xi z \quad (1.3)$$

and to write $\Lambda_\theta(z)$ as $\tilde{\Lambda}_\theta(v(z))$. Up to transformation $v \mapsto 1 - v$, this is just the quantile transformation, i.e.; the distribution of $\mathcal{L}(\Lambda_\theta(\frac{X-\mu}{\beta}))$ for $X \sim \text{GPD}(\theta)$ is just $\mathcal{L}(\tilde{\Lambda}_\theta(V))$ for $V \sim \text{unif}(0, 1)$.

Using transformation (1.3), we easily obtain Fisher information \mathcal{I}_θ as

$$\mathcal{I}_\theta = \frac{1}{(2\xi + 1)(\xi + 1)} \begin{pmatrix} 2, & \beta^{-1} \\ \beta^{-1}, & \beta^{-2}(\xi + 1) \end{pmatrix} \quad (1.4)$$

As \mathcal{I}_θ is positive definite for $\xi > 0, \beta > 0$, the model is (locally) identifiable.

The model also is *scale invariant*. Using matrix $d_\beta = \text{diag}(1, \beta)$. Correspondingly, an estimator S for $\theta = (\xi, \beta)$ is called *(scale)-equivariant* if

$$S(\beta x_1, \dots, \beta x_n) = d_\beta S(x_1, \dots, x_n) \quad (1.5)$$

and in terms of the L_2 derivative, we have

$$\Lambda_{(\xi, \beta)}(z) = d_\beta^{-1} \Lambda_{(\xi, 1)}(z) \quad (1.6)$$

To preserve this invariance when determining the ‘‘length’’ of a parameter, Robust Statistics uses special norms for the parameter space; as a simple scale invariant norm, we use the weighted norm

$$n_\beta(x, y) = \|d_\beta^{-1}(x, y)\| = \sqrt{x^2 + y^2/\beta^2} \quad (1.7)$$

Remark 1.2 For the shape parameter there is no obvious such invariance, except for the quantile transformation, of course, i.e.;

$$g(\theta, \theta'; x) = F_{\theta'}^{-1} \circ F_\theta(x) = [(1 + \xi x/\beta)^{\xi'/\xi} - 1] \beta'/\xi' \quad (1.8)$$

transforming an F_θ -distributed observation X into an $F_{\theta'}$ -distributed one. The only values of x invariant under arbitrary $g(\theta, \theta'; \cdot)$ are $\{0, \infty\}$, as in the pure scale case. However, with this group, we do not see any form of reasonable equivariance.

Gross Error Model Instead of working only with ideal distributions, Robust Statistics considers suitable distributional neighborhoods about this ideal model. In this paper, we limit ourselves to the *Gross Error Model*, i.e.; our neighborhoods are the sets of all distributions F^{re} representable as

$$F^{\text{re}} = (1 - \varepsilon)F^{\text{id}} + \varepsilon F^{\text{di}} \quad (1.9)$$

for some given size or radius $\varepsilon > 0$, where F^{id} is the underlying ideal distribution and F^{di} some arbitrary, unknown, and uncontrollable contaminating distribution. For fixed $\varepsilon > 0$, bias and variance scale at different rates ($O(1)$, $O(1/n)$, resp.). Hence to balance these scales, in the shrinking neighborhood approach as developed (a.o.) in Huber-Carol (1970), Rieder (1994, 1978), and Bickel (1981), one lets the radius of these neighborhoods shrink with growing sample size n , i.e.;

$$\varepsilon = r_n = r/\sqrt{n} \quad (1.10)$$

(and the contamination G may vary in n as well).

In reality one rarely knows ε or r , but Rieder et al. (2008) give objective criteria for their choice to specify a procedure in situations where one has no or only limited knowledge of the “true” radius. For our numerical and simulational evaluations, we use a starting radius $r = 0.5$.

Remark 1.3 $r = 0.5$ is very close to the minimax radius in the situation where we have no knowledge at all about the radius, which for $\xi = 0.7, \beta = 1$ would be 0.486, leading to a maximin efficiency of 0.683, i.e.; using the resp. radius minimax procedure, the performance of this procedure would never be worse than 1.464 times the maximal asMSE (see below) of the optimal procedure knowing the radius. The minimal efficiency of the OMSE to radius $r = 0.5$ is in fact only 0.678 (achieved when used for unknown radius $r = 0$).

1.2 Robustness

Robustness distinguishes local properties (measuring the infinitesimal influence of a single observation) like the *influence function* (IF) and global ones (measuring the effect of massive deviations) like the *breakdown point*.

Influence Function Defining an estimator as a functional T evaluated at the empirical distribution, the IF of T is the functional derivative of the estimator with respect to the distribution. Historically, in Hampel (1968) this is defined as the Gâteaux derivative in the direction of a Dirac measure δ_x (provided the limit exists): For $F_\varepsilon = (1 - \varepsilon)F + \varepsilon\delta_x$ and F the underlying distribution, the influence function (IF) of the estimator T at x then is

$$\text{IF}(x; T, F) = \lim_{\varepsilon \rightarrow 0} \frac{T(F_\varepsilon) - T(F)}{\varepsilon} \quad (1.11)$$

This definition however has its flaws (Kohl et al., 2010, introduction). Fortunately, using the (finite-dim.) Delta method, in our context, everything can be reduced to the question of differentiability of the likelihood (MLE, SMLE), of quantiles (PE, PE*, PicM, MMed, MedMAD, kMedMAD), and of the c.d.f. (MDE), so these gaps can be closed by results from Fernholz (1979), Rieder (1994, Chap. 1) for our estimators; the results on one-step estimators of Rieder (1994, Chap. 6) show that MBRE and OMSE do have influence functions.

Remark 1.4 Assuming an L_2 -differentiable model, for our purposes, we need the property that the estimator S_n has the expansion in the observations X_i as

$$S_n = \theta + \frac{1}{n} \sum_{i=1}^n \psi_\theta(X_i) + R_n, \quad \sqrt{n}|R_n| \xrightarrow{n \rightarrow \infty} 0 \quad P_\theta^n\text{-stoch.} \quad (1.12)$$

for $\psi_\theta \in L_2(P_\theta)$ the IF of S_n for which we require

$$E_\theta \psi_\theta = 0, \quad E_\theta \psi_\theta \Lambda_\theta^\tau = \mathbb{I}_k \quad (1.13)$$

Equation (1.13) may be motivated either by Rieder (1994, Lemma 4.2.18) or R. and H. (2010a, Lemma 1.3). An estimator with (1.12) is called *asymptotically linear* or *ALE*. We note that all estimators considered in this paper are ALEs.

In shrinking neighborhood approach, if well initialized, all relevant asymptotic properties of an ALE except for its breakdown point only depend on its IF:

Asymptotic Variance The asymptotic (co)variance matrix ASV of an ALE S_n may be determined as

$$\text{asVar}(S_n) = \int \psi_\theta \psi_\theta^\tau dF_\theta \quad (1.14)$$

(Rieder, 1994, Rem. 4.2.17(b)).

Asymptotic Bias The *gross error sensitivity* GES, see Hampel et al. (1986, Chapter 2.1c), is defined as

$$\text{GES} := \sup_x |\psi_\theta(x)| \quad (1.15)$$

It may be shown (Rieder, 1994, Lemma 5.3.3), that in the shrinking neighborhood setup, the \sqrt{n} -standardized, maximal asymptotic bias of an ALE S_n in the gross error model (1.9), (1.10) is just

$$\text{asBias}(S_n) = r \text{GES} = r \sup_x |\psi_\theta(x)| \quad (1.16)$$

Asymptotic MSE As a consequence of the previous two paragraphs, the (maximal, standardized) asymptotic mean squared error (MSE) attainable in the gross error model (1.9), (1.10) with starting radius r can be calculated as

$$\text{asMSE}(S_n) = r^2 \text{GES}^2 + \text{tr}(\text{asVar}(S_n)) \quad (1.17)$$

Suitable constructions (Rieder, 1994, chap. 6) allow to interchange quantors and asMSE also is the standardized asymptotic maximal MSE.

Remark 1.5 It is common in Robust Statistics to use high breakdown point estimators (see below) tuned to a high efficiency (say 95%) in the ideal model in a reweighting step. But efficiency in the ideal model is a bad scale in the presence of outliers, as the “insurance premium” paid in terms of the 5% efficiency loss does not reflect the protection “bought”, as this protection will vary from model to model, and in our non-invariant case even from θ to θ . Instead, we prefer the minimax criteria asMSE, asBias on whole neighborhoods to define optimally robust estimators (OMSE, MBRE). Illustrating this, the OBRE tuned for 95% efficiency in the ideal model at $\xi = 0.7$ may drop down to 14% efficiency for sufficiently large radius (in comparison to the best procedure knowing the radius), while OMSE never falls below 68% no matter what radius.

Efficiency We also determine efficiencies of estimators S_n in the ideal model (eff.id) and under contamination of radius $r = 0.5$ (eff.re),

$$\text{eff.id}(S_n) = \frac{\text{tr}(\mathcal{J}^{-1})}{\text{tr}(\text{asVar}(S_n))}, \quad \text{eff.re}(S_n) = \frac{\text{asMSE}(\text{OMSE})}{\text{asMSE}(S_n)} \quad (1.18)$$

In addition, for the situation where radius r is unknown, we also compute the least favorable efficiency of each (fixed) estimator (i.e.; we still “guess” that $r = 0.5$ for OMSE, although this is presumably false) w.r.t. the most efficient procedure knowing the radius, denoted by eff.ru. For this notion, see Rieder et al. (2008). These efficiencies may be read as the relative amount of observations, the optimal procedure (MLE in the ideal setting, $\text{OMSE}_{r=0.5}$ under contamination of known radius $r = 0.5$, and $\text{OMSE}_{r=r_{1.f.}}$ for least favorable actual radius $r_{1.f.}$ for contamination of unknown radius) would need to achieve the same accuracy as the estimator under consideration. As in Kohl (2005, Lemma 2.2.3(a)), we see that for all considered estimators S_n

$$\text{eff.ru}(S_n) = \min(\text{eff.id}(S_n), \text{GES}^2(\text{MBRE})/\text{GES}^2(S_n)) \quad (1.19)$$

Thus the least favorable (unknown) radius is either $r = 0$ or $r = \infty$, and, to be precise, for all estimators but kMedMAD and MBRE, it is $r = \infty$.

Breakdown Point The breakdown point in the gross error model (1.9) gives the largest radius ε at which the estimator still produces reliable results. We take the definitions from Hampel et al. (1986, 2.2 Definitions 1,2):

The *asymptotic breakdown point (ABP)* ε^* of the sequence of estimators T_n for parameter $\theta \in \Theta$ at probability F is given by

$$\varepsilon^* := \sup \left\{ \varepsilon \in (0, 1]; \text{ there is a compact set } K_\varepsilon \subset \Theta \text{ s.t.} \right. \\ \left. \pi(F, G) < \varepsilon \implies G(\{T_n \in K_\varepsilon\}) \xrightarrow{n \rightarrow \infty} 1 \right\}, \quad (1.20)$$

where π is Prokhorov distance.

The *finite sample breakdown point (FSBP)* ε_n^* of the estimator T_n at the sample (x_1, \dots, x_n) is given by

$$\varepsilon_n^*(T_n; x_1, \dots, x_n) := \frac{1}{n} \max \left\{ m; \max_{i_1, \dots, i_m} \sup_{y_1, \dots, y_m} |T_n(z_1, \dots, z_n)| < \infty \right\}, \quad (1.21)$$

where the sample (z_1, \dots, z_n) is obtained by replacing the data points x_{i_1}, \dots, x_{i_m} by arbitrary values y_1, \dots, y_m . The ABP was introduced in Hampel (1968), and the FSBP in Donoho and Huber (1983). For deciding upon which procedure to take *before* having made observations, in particular for ranking procedures in a simulation study, the FSBP from (1.21) has some drawbacks: for some of the considered estimators, the dependence on possibly highly improbable configurations of the sample entails that not even a non-trivial lower bound for the FSBP exists. To get rid of this dependence to some extent at least, but still preserving the aspect of a finite sample, we hence use the *expected FSBP* as proposed and worked out to some detail in R. and H. (2010b), i.e.;

$$\bar{\varepsilon}_n^*(T_n) := \mathbb{E} \varepsilon_n^*(T_n; X_1, \dots, X_n) \quad (1.22)$$

where expectation is evaluated in the ideal model. We also consider the limit $\bar{\varepsilon}^*(T) := \lim_{n \rightarrow \infty} \bar{\varepsilon}_n^*(T_n)$ and also call it EFSBP where unambiguous.

1.3 Computational and Numerical Aspects

For an estimator to be useful in practice also computational aspects deserve attention. In this respect, our estimator can be divided into four classes:

1. Estimators in closed-form expressions like PE (after possibly sorting the observations). As to computation time, their evaluation is by magnitudes faster than of the other groups, which makes them attractive for batch uses.

2. M-estimators like MLE, SMLE, and MDE, obtained by optimizing a corresponding criterion function and solved iteratively by using R function `optim` and hence need a suitable initialization to find the “right” local optimum.

3. Z-estimators like MMed and kMedMAD, i.e.; the zero of $a(n)$ (system of equation(s)). In fact, both cases may be reduced to univariate problems, hence may use R function `uniroot`, with canonical search interval.

4. One-step constructions like MBRE and OMSE, depending on a suitably chosen starting estimator. Once this starting estimate is found and the respective influence function at the starting estimate determined, computation of MBRE and OMSE is extremely fast, just involving an average. The computation of the influence function at the starting estimate is not trivial, however, and to speed this up, we present Algorithm 2.8.

For computations, we use R, R Development Core Team (2009), and addon-packages `R0ptEst`, Kohl and R. (2009), `POT`, Ribatet (2009), available on `cran.r-project.org`.

2 Estimators

2.1 Maximum Likelihood Estimator

The maximum likelihood estimator is the maximizer (in θ) of the (product-log-) likelihood $l_n(\theta; X_1, \dots, X_n)$ of our model

$$l_n(\theta; X_1, \dots, X_n) = \sum_{i=1}^n l_\theta(X_i), \quad l_\theta(x) = \log f_\theta(x) \quad (2.1)$$

For the GPD, this maximizer has no closed-form solutions and has to be determined numerically, using a suitable initialization; in our simulation study, we use the Hybr estimator as defined in Subsection 2.6.

IF The MLE admits as influence function

$$\text{IF}_\theta(z; \text{MLE}, F) = \mathcal{J}_\theta^{-1} \Lambda_\theta(z) \quad (2.2)$$

Regularity conditions, e.g. van der Vaart (1998, Thm. 5.39), can easily be checked due to the smoothness of the scores function. In particular, MLE attains the smallest asymptotic variance among all ALEs according to the Asymptotic Minimax Theorem, Rieder (1994, Thm. 3.3.8). Using the quantile-type representation (1.3), we obtain

$$\tilde{\psi}(v) = \frac{\xi+1}{\xi^2} \begin{pmatrix} -(\xi^2 + \xi) \log(v) + (2\xi^2 + 3\xi + 1)v^\xi - (\xi^2 + 3\xi + 1) \\ \xi \log(v) - (2\xi^2 + 3\xi + 1)v^\xi + (3\xi + 1) \end{pmatrix} \quad (2.3)$$

As to in-/equivariance, we note that

$$\text{IF}_{(\xi,\beta)}(x; \text{MLE}, F) = d_\beta \text{IF}_{(\xi,1)}(x/\beta; \text{MLE}, F) \quad (2.4)$$

hence, as MLE is an ALE, we have asymptotic equivariance by (1.12).

ASV The asymptotic covariance matrix of the maximum likelihood estimators is equal to the inverse of the Fisher information function:

$$\mathcal{I}_\theta^{-1} = (1 + \xi) \begin{pmatrix} \xi + 1, & -\beta \\ -\beta, & 2\beta^2 \end{pmatrix} \quad (2.5)$$

ASB As $(\mathcal{I}_\theta^{-1})_{1,1}, (\mathcal{I}_\theta^{-1})_{2,1} \neq 0$, both components of the influence curve are unbounded (although only growing in absolute value at rate $\log(x)$). Hence, for any neighborhood of positive radius, we can induce arbitrarily large bias, so MLE is not robust.

FSBP By standard arguments, MLE is shown to have a FSBP of $1/n$, i.e.; arbitrarily close to 0 for large n . Admittedly, though, one only can approximate this breakdown for finite samples and finite contamination with really large contaminations.

2.2 Skipped Maximum Likelihood Estimators

Skipped Maximum Likelihood Estimators (SMLE) are ordinary MLE, skipping the largest k observations. This has to be distinguished from the better investigated *trimmed/weighted MLE*, studied by Field and Smith (1994), Hadi and Luceño (1997), Vandev and Neykov (1998), Müller and Neykov (2001), where trimming/weighting is done according to the size (in absolute value) of the log-likelihood.

In general these concepts fall apart as they refer to different orderings; in our situation though they coincide due to the monotonicity of the likelihood in the observations.

As this skipping is not done symmetrically, it induces a non-vanishing bias $B_n = B_{n,\theta}$ already present in the ideal model. To cope with such biases three strategies can be used—the first two already considered in detail in Dupuis and Morgenthaler (2002, Section 2.2): (1) correcting the criterion function for the skipped summands, (2) correcting the estimator for bias B_n , and (3) no bias correction at all, but, conformal to our shrinking neighborhood setting, to let the skipping proportion α shrink at the same rate. Strategy (3) reflects the common practice where α is often chosen small, and the bias correction is omitted. In the sequel, we only study Strategy (3) with $\alpha = \alpha_n = r'/\sqrt{n}$ for some r' larger than the actual r . This way indeed bias becomes asymptotically negligible:

Lemma 2.1 *In our ideal GPD model, eventually in n , the bias B_n of SMLE with skipping rate α_n is bounded from above by $\bar{c}\alpha_n \log(n)$ for some $\bar{c} < \infty$.*

If for some $0 < \beta \leq 1$, $\liminf_n \alpha_n n^\beta > 0$, then also $\liminf_n n^\beta B_n \geq \underline{c} \liminf_n n^\beta \alpha_n \log(n)$ for some $\underline{c} > 0$.

If $0 < \underline{\alpha} = \liminf_n \alpha_n < \alpha_0$ for $\alpha_0 = \exp(-3 - 1/\xi)$, then $\liminf_n B_n \geq \underline{c}' \underline{\alpha} (-\log(\underline{\alpha}))$ for some $\underline{c}' > 0$.

A proof to this Lemma can be found in R. and H. (2010a).

Hence, for higher FSBPs, we need to correct for the then considerable bias. Obviously this can cope with $\alpha_n n$ outliers.

IF As we have seen, SMLE in fact does not estimate θ but $d(\theta) = \theta + B_\theta$, for the bias B_θ already present in the ideal model. So to determine the IF for this estimator, we only compute the influence function for the functional estimating $d(\theta)$. To this end, we may use the underlying order statistics of the X_i and obtain the IF of SMLE just as the IF of the L-estimate to the following functional:

$$T(F) = \frac{1}{1-\alpha} \int_0^{1-\alpha} \Lambda_\theta(F^{-1}(s)) ds \quad (2.6)$$

The influence function, referring to Huber (1981, Chapter 3.3), is analogous to the influence function of the trimmed mean (with $u_\alpha := F^{-1}(1-\alpha)$):

$$\text{IF}_\theta(z; \text{SMLE}, F) = \mathcal{J}_\theta^{-1} \begin{cases} \frac{1}{1-\alpha} [\Lambda_\theta(z) - W(F)], & 0 \leq x \leq u_\alpha \\ \frac{1}{1-\alpha} [\Lambda_\theta(u_\alpha) - W(F)], & x > u_\alpha \end{cases} \quad (2.7)$$

$$W(F) = (1-\alpha) \text{SMLE}(F) + \alpha \Lambda_\theta(u_\alpha) \quad (2.8)$$

It enjoys the same (asympt.) equivariance (1.5) as the MLE.

ASV Analytic terms of the asymptotic covariance of the SMLE are not available; instead we only include numerical values in the tables in Section 3.

ASB By Lemma 2.1, for a shrinking rate $\alpha_n = r'/\sqrt{n}$, asymptotic bias of SMLE is finite for each n , but, standardized by \sqrt{n} , is of order $\log(n)$, hence unbounded. As the IF is bounded locally uniform in θ , the extra bias induced by contamination is dominated by B_n eventually.

FSBP In our shrinking setting the proportion of the skipped data tends to 0, so it is this proportion which delivers the active bound for the breakdown point: Just replace $\lceil \alpha_n n \rceil + 1$ observations by something sufficiently large and argue as for the MLE to show that $\text{FSBP} = \alpha_n$.

2.3 Cramér-von-Mises Minimum Distance Estimators

General minimum distance estimators are defined as minimizers of a suitable distance between the theoretical F and empirical distribution \hat{F}_n . Optimization of this distance in general has to be done numerically and, as for MLE and SMLE, depends on a suitable initialization. We use Cramér-von-Mises distance defined for c.d.f.'s F, G and some σ -finite measure ν on \mathbb{B}^k as

$$d_{\text{cvm}}(F, G)^2 = \int (F(x) - G(x))^2 \nu(dx) \quad (2.9)$$

i.e.; by MDE we denote

$$\text{MDE} = \text{argmin}_\theta d_{\text{cvm}}(\hat{F}_n, F_\theta) \quad (2.10)$$

In this paper, we use¹ $\nu = P_\theta$. Hybr from Subsection 2.6 again serves as initialization. MDE is known to have good global robustness properties: it is an ALE with bounded

¹ Another setting common in the literature uses the empirical, $\nu = \hat{F}_n$.

IF (Rieder, 1994, Rem 6.3.9(a), 4.2 eq.(55)) and, according to Donoho and Liu (1988), up to factor 2 achieves the smallest sensitivity to contamination among Fisher-consistent estimators.

IF For the influence function of MDE, we follow Rieder (1994, Example 4.2.15, Theorem 6.3.8) and obtain

$$\text{IF}(x; \text{MDE}, F) =: \mathcal{J}_\theta^{-1}(\tilde{\varphi}_\xi(x), \tilde{\varphi}_\beta(x)) \quad (2.11)$$

where for v from (1.3) it holds that

$$\tilde{\varphi}_\xi(v(z)) = \frac{19+5\xi}{36(3+\xi)(2+\xi)} + \frac{1}{\xi}v^2 \log(v) + \frac{2-\xi}{4\xi^2}v^2 - \frac{1}{\xi^2(2+\xi)}v^{2+\xi} \quad (2.12)$$

$$\tilde{\varphi}_\beta(v(z)) = \frac{5+\xi}{6(3+\xi)(2+\xi)\beta} - \frac{1}{2\xi\beta}v^2 + \frac{1}{\xi\beta(2+\xi)}v^{2+\xi} \quad (2.13)$$

and \mathcal{J}_θ is the CvM Fisher information as defined, e.g. in Rieder (1994, Definition 2.3.11)). We have

$$\mathcal{J}_\theta^{-1} = 3(\xi+3)^2 \begin{pmatrix} \frac{18(\xi+3)}{(2\xi+9)}, & -3\beta \\ -3\beta, & 2\beta^2 \end{pmatrix} \quad (2.14)$$

Apparently the same (asympt.) in-/equivariance as for MLE and SMLE holds again.

Remark 2.2 The fact that MDE is asymptotically linear with the IF just given allows for an alternative to the numerical minimization of the distance: we could instead use a corresponding one-step construction built up on a suitable starting estimator. Asymptotically both variants will be indistinguishable.

ASV The asymptotic covariance of the CvM minimum distance estimators can be found analytically or numerically. Its analytic terms² are rational functions in ξ and β :

$$\text{as Var}(\text{MDE}) = \frac{(3+\xi)^2}{125(5+2\xi)(5+\xi)^2} \begin{pmatrix} V_{1,1}, & V_{1,2} \\ V_{1,2}, & V_{2,2} \end{pmatrix} \quad (2.15)$$

for

$$V_{1,1} = 81 \left(16\xi^5 + 272\xi^4 + 1694\xi^3 + 4853\xi^2 + 7276\xi + 6245 \right) (2\xi+9)^{-2}, \quad (2.16)$$

$$V_{1,2} = -9\beta \left(4\xi^4 + 86\xi^3 + 648\xi^2 + 2623\xi + 4535 \right) (2\xi+9)^{-1}, \quad (2.17)$$

$$V_{2,2} = \beta^2 \left(26\xi^3 + 601\xi^2 + 3154\xi + 5255 \right) \quad (2.18)$$

ASB As noted, the IF of MDE is known to be bounded, so ASB is finite.

FSBP Due to the lack of invariance in the GPD situation, Donoho and Liu (1988, Propositions 4.1 and 6.4) only provide bounds for the FSBP, telling us that its FSBP must be no smaller than 1/2 the FSBP of the FSBP-optimal procedure. As MDE is a minimum of the smooth CvM distance, it has to fulfill the first order condition for the corresponding M-equation, i.e.; for $V_i = \left(1 + \frac{\xi}{\beta}X_i\right)^{-1/\xi}$,

$$\sum_i \tilde{\varphi}_\xi(V_i; \xi) = 0, \quad \sum_i \tilde{\varphi}_\beta(V_i; \xi) = 0 \quad (2.19)$$

² For the interested reader willing to control these formula, we have MAPLE scripts to determine them.

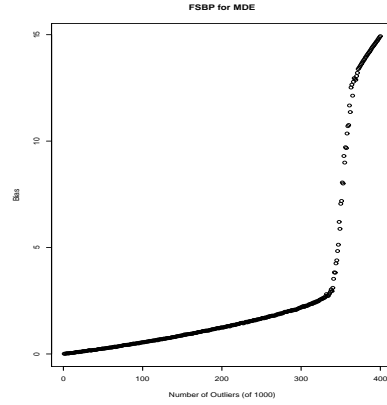


Fig. 1 Empirical Bias for FSBP of MDE CvM

Arguing as for the breakdown point of an M-estimator, except for the optimization in ξ , we obtain the following analogue to Huber (1981, Chap. 3, eqs. (2.39) and (2.40)):

$$\varepsilon_n^* \leq \min \left\{ \frac{-\inf_{v,\xi} \varphi}{\sup_{v,\xi} \varphi - \inf_{v,\xi} \varphi}, \frac{\sup_{v,\xi} \varphi}{\sup_{v,\xi} \varphi - \inf_{v,\xi} \varphi}, \cdot = \xi, \beta \right\} \quad (2.20)$$

although, to make the inequality in (2.20) an equality, we would need to show that we cannot produce a breakdown with less than this bound. Evaluating bound (2.20) numerically gives a value of $4/9 \doteq 36.37\%$, which is achieved for $v = 0$ (and $\xi \rightarrow 0$) or, equivalently, letting the m replacing observations in Definition 1.2 tend to ∞ .

Remark 2.3 To see how realistic this value is, in Figure 1, we produce an empirical max-bias-curve, simulating $M = 100$ samples of size $n = 1000$ observations from a GPD with $\xi = 0.7$, $\beta = 1$, and after replacing m observations, for $m = 1, \dots, 400$ by value 10^{10} compute the bias. There is a steep increase around 354, so we conjecture that (E)FSBP should be approximately 0.35; on the other side, MDE cannot have a higher FSBP than its initialization, and so far the best known initialization has (E)FSBP of 0.346.

2.4 Pickands Estimator

Estimators based on the empirical quantiles of GPD are described in the Elementary Percentile Method (EPM) by Castillo and Hadi (1997). Pickands estimator (PE), a special case of EPM, is based on the empirical 50% and 75% quantiles M_2 and M_4 respectively, and has first been proposed by Pickands (1975). The construction behind PE is not limited to 50% and 75% quantiles. More specifically, let $a > 1$ and consider the empirical α_i -quantiles for $\alpha_1 = 1 - 1/a$ and $\alpha_2 = 1 - 1/a^2$ denoted by $M_2(a)$, $M_4(a)$, respectively. Then PE is obtained for $a = 2$, and as theoretical quantiles we obtain $M_2^{\natural}(a) = \frac{\beta}{\xi}(a^\xi - 1)$, $M_4^{\natural}(a) = \frac{\beta}{\xi}(a^{2\xi} - 1)$, and the (generalized) PE denoted by PE(a) for ξ and β is

$$\hat{\xi} = \frac{1}{\log a} \log \frac{M_4(a) - M_2(a)}{M_2(a)}, \quad \hat{\beta} = \hat{\xi} \frac{M_2(a)^2}{M_4(a) - 2M_2(a)} \quad (2.21)$$

Apparently for any $a > 1$, PE(a) enjoys the corresponding equivariance as MLE, SMLE, and MDE.

IF The influence function of linear combinations T_L of the quantile functionals $F^{-1}(\alpha_i) = T_i(F)$ for probabilities α_i and weights h_i , $i = 1, \dots, k$ may be read off from Rieder (1994, Chapter 1.5) and gives

$$\text{IF}(x; T_L, F) = \sum_{i=1}^k h_i (\alpha_i - \mathbb{I}(x \leq F^{-1}(\alpha_i))) / f(F^{-1}(\alpha_i)) \quad (2.22)$$

Using the Δ -method, the influence functions of PE(a) hence is

$$\text{IF}_\bullet(x; \text{PE}(a), F) = \sum_{i=1,2} h_{\bullet,i}(a) \frac{\alpha_i(a) - \mathbb{I}(x \leq M_{2i}(a))}{f(M_{2i}(a))}, \quad \bullet = \xi, \beta \quad (2.23)$$

with weights $h_{\bullet,i}(a)$ to be read off from R. and H. (2010a, eqs.(2.43)-(2.45)) Apparently we have again (asympt.) equivariance,

$$\text{IF}_{(\xi, \beta)}(x; \text{PE}(a), F) = d_\beta \text{IF}_{(\xi, 1)}(x/\beta; \text{PE}(a), F) \quad (2.24)$$

ASV Abbreviating $\alpha_i(a)$ by α_i , $1 - \alpha_i$ by $\bar{\alpha}_i$, and $h_{\bullet,1}(a)$ by $h_{\bullet,1}$, $\bullet = \xi, \beta$, the asymptotic covariance for PE(a) is

$$\text{as Var}(\text{PE}(a)) = D(a)^T \Sigma(a) D(a), \quad (2.25)$$

$$\Sigma(a) = \beta^2 \begin{pmatrix} \alpha_1 \bar{\alpha}_1^{-1-2\xi} & \alpha_1 \bar{\alpha}_1^{-1-\xi} \bar{\alpha}_2^{-\xi} \\ \alpha_1 \bar{\alpha}_1^{-1-\xi} \bar{\alpha}_2^{-\xi} & \alpha_2 \bar{\alpha}_2^{-1-2\xi} \end{pmatrix}, \quad D(a) = \begin{pmatrix} h_{\xi,1} & h_{\xi,2} \\ h_{\beta,1} & h_{\beta,2} \end{pmatrix} \quad (2.26)$$

ASB The IF of PE(a) is bounded, so ASB is finite.

FSBP With simple generalizations we may refer to R. and H. (2010b) to show that

$$\varepsilon_n^* = \min\{1/a^2, \hat{N}^0/n\}, \quad \hat{N}_n^0 := \#\{X_i \mid 2M_2(a) \leq X_i \leq M_4(a)\} \quad (2.27)$$

By usual LLN arguments, $\hat{N}^0/n \rightarrow \pi_\xi(a) = (2a^\xi - 1)^{-1/\xi} - 1/a^2$, so that

$$\bar{\varepsilon}^* = \bar{\varepsilon}^*(a) = \min\{\pi_\xi(a), 1/a^2\} \quad (2.28)$$

For $\xi = 0.7$, the classical PE achieves an ABP of $\bar{\varepsilon}^*(a=2) \doteq 6.42\%$; as to EFSBP, for $n = 40, 100, 1000$ we obtain $\bar{\varepsilon}_n^* = 9.48\%, 7.61\%, 6.53\%$, respectively (R. and H., 2010b, Table 2).

Remark 2.4 Optimizing for a high (E)FSBP within the class of PE(a) estimator, one obtains estimator PE* (R. and H., 2010a), which in case of our reference parameter $\xi = 0.7$ gives $a^* = 2.658$ with a EFSBP of 7.02%, so we have not won much. Similarly, tuning for a better variance by averaging several PE(a)'s for varying a (PicM in the cited reference) does improve the efficiencies, but still does not give convincing results.

2.5 Method of Medians Estimator

The Method of Medians estimator of Peng and Welsch (2001) consists in fitting the (population) medians of the the two coordinates of the scores function Λ_θ against the corresponding sample medians, i.e.; we have to solve the system of equations

$$\text{Median}(X_i)/\beta = F_{1,\xi}^{-1}(1/2) = (2^\xi - 1)/\xi =: m_\xi \quad (2.29)$$

$$\text{Median}\left(\log(1 + \xi X_i/\beta)\beta^{-2} - (1 + \xi)X_i(\beta\xi + \xi^2 X_i)^{-1}\right) = z(\xi) \quad (2.30)$$

where $z(\xi)$ is the population median of the ξ -coordinate of $\Lambda_{(1,\xi)}(X)$, $\Lambda_{(1,\xi);2}(X)$ for $X \sim \text{GPD}(1, \xi)$. Solving the first equation for β and plugging in the corresponding expression in ξ into the second equation, we obtain a one-dimensional root-finding problem to be solved, e.g. in R by `uniroot`. In the same sense as the estimators considered so far, the MMed is equivariant.

IF The IF of MMed is a linear combination of the IF of the sample median already used for the PE, and the IF of the median of the ξ -coordinate of $\Lambda_{(1,\xi);2}(X)$. Now, as can be seen when plotting the function $x \mapsto \Lambda_{(1,\xi);2}(x)$, for $\xi = 0.7$, the level set $\Lambda_{(1,\xi);2}(X) \leq z(\xi)$ is of form $[q_1(\xi), q_2(\xi)]$, so that

$$\text{IF}(x; \Lambda\text{-Med}, F) = \frac{\mathbb{I}(q_1 \leq x \leq q_2) - 1/2}{f_\theta(q_2)/l_2 - f_\theta(q_1)/l_1} \quad (2.31)$$

where $l_i := \frac{\partial}{\partial x} \Lambda_{(1,\xi);2}(q_i)$. More precisely, for $\xi = 0.7$, we obtain $q_1 \doteq 0.3457$ and $q_2 \doteq 2.5449$. In analogy to the Pickands-type estimators we could now determine a corresponding Jacobian D in closed form such that

$$\text{IF}(x; \text{MMed}, F) = D(\text{IF}(x; \text{Median}, F), \text{IF}(x; \Lambda\text{-Med}, F))^\tau \quad (2.32)$$

but in our context it is easier to determine \tilde{D} numerically by

$$\tilde{D}^{-1} = E_\theta \eta_\theta \Lambda_\theta^\tau \text{ for } \eta_\theta(x) = \left(\mathbb{I}(x \leq m_\xi) - 1/2, \mathbb{I}(q_1 \leq x \leq q_2) - 1/2 \right)^\tau \quad (2.33)$$

and then to write

$$\text{IF}(x; \text{MMed}, F) = \tilde{D} \eta_\theta \quad (2.34)$$

Corresponding analytic terms may be found in Peng and Welsch (2001, p. 60).

ASV Similarly, we obtain

$$\text{asVar}(\text{MMed}) = \tilde{D} \Sigma(a) \tilde{D}^\tau, \quad \Sigma(a) = \frac{1}{4} \begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix}, \quad c = 1 - 4F(q_1) \quad (2.35)$$

ASB The IF of MMed is bounded, so ASB is finite.

FSBP We have not found analytic values for neither the asymptotic nor the finite sample breakdown point. While 50% by equivariance is an upper bound, the high frequency of failures in the simulation study for small sample sizes however indicates that (E)FSBP should be considerably smaller; a similar study for the empirical maxBias as the one for MDE gives that for sample size n from a rate of outliers of ε_n on, we have but failures in solving for MMed, for $\varepsilon_{40} = 42.5\%$, $\varepsilon_{100} = 35.0\%$, $\varepsilon_{1000} = 25.1\%$, and $\varepsilon_{10000} = 20.1\%$. So we conjecture that the asymptotic breakdown point $\varepsilon^* \leq 20\%$.

2.6 kMedMAD

Empirical median $\hat{m} = \hat{m}_n$ and median of absolute deviations (MAD) $\hat{M} = \hat{M}_n$ are well known for their high breakdown point, jointly attaining the highest possible ABP of 50% among all affine equivariant estimators at symmetric, continuous univariate distributions.

Hence it is plausible to define an estimator for ξ and β , matching \hat{m} and \hat{M} against their population counterparts m and M within the GPD model. Now it turns out that the mapping $(\xi, \beta) \mapsto (m, M)(F_\theta)$ is indeed a diffeomorphism, hence we can solve the implicit equations for ξ, β to obtain an estimator introduced as *MedMAD* in R. and H. (2010a).

Due to the considerable skewness to the right of the GPD, this estimator can be improved though by using a scale estimator that takes this skewness into account: For a distribution F on \mathbb{R} with median m let us define for $k > 0$

$$\text{kMAD}(F, k) := \inf \{ t > 0 \mid F(m + kt) - F(m - t) \geq 1/2 \} \quad (2.36)$$

where k in our case is chosen to be a suitable number larger than 1, and $k = 1$ would reproduce the MAD; i.e.; within the class of intervals about the median m with covering probability 50%, we only search those where the part right to m is k times longer than the one left to m . Whenever F is continuous, kMAD preserves the FSBP of the MAD of 50%.

The corresponding estimator for ξ and β is called *kMedMAD* and consists of two estimating equations. The first equation is for the median of the GPD, which is $m = m(\xi, \beta) = F^{-1}(0.5) = \beta(2^\xi - 1)/\xi$. The second equation is for the respective kMAD, which has to be solved numerically as unique root M of $f_{m, \xi, \beta; k}(M)$ for

$$f_{m, \xi, \beta; k}(M) = -v_+ + v_- - \frac{1}{2} \quad (2.37)$$

where

$$v_+ := \left(1 + \xi \frac{kM+m}{\beta}\right)^{-\frac{1}{\xi}}, \quad v_- := \left(1 + \xi \frac{m-M}{\beta}\right)^{-\frac{1}{\xi}} \quad (2.38)$$

Note that for any distribution G on \mathbb{R} with $G((-\infty; p]) = 0$ for some finite p , and any $k > 0$, $\text{kMAD}(G; k) \leq \text{median}(G)$ with equality if and only if $G(\{\text{median}(G)\}) \geq 0.5$. Consequently, $f_{m, \xi, \beta; k}(M) > 0$ for $M \geq m$, hence the population kMAD $M_k(\xi, \beta)$ in the GPD must always be smaller than its median, or $M_k(\xi, \beta)/m(\xi, \beta) < 1$.

Now, kMAD is scale-invariant, i.e.; $M_k(\xi, \beta) = \beta M_k(\xi, 1)$, and the empirical kMAD \hat{M}_k is scale-equivariant, i.e.; $\hat{M}_k(\beta x_1, \dots, \beta x_n) = \beta \hat{M}_k(x_1, \dots, x_n)$. The same in-/equivariance also holds for the median; hence the quotient $q_k(\xi) := M_k(\xi, \beta)/m(\xi, \beta)$ and its empirical counterpart $\hat{q}_{k;n}$ are scale-free; so we have reduced the problem by one dimension.

In R. and H. (2010b), plotting for given k the function $\xi \mapsto q_k(\xi)$, one sees that q_k is strictly isotone, but that there is a second restriction of the same sort as that $q_k(\xi) < 1$, induced by the fact that for all $\xi > 0$

$$q_k(\xi) \geq \lim_{\xi \rightarrow 0} q_k(\xi) =: \check{q}_k \quad (2.39)$$

Hence matching $\hat{q}_{k;n}$ against $q_k(\xi)$ amounts to finding a zero $\hat{\xi}_n$ of $G(\xi) = q_k(\xi) - \hat{q}_{k;n}$ in the interval $(\check{q}_k; 1)$ which can easily be solved with a standard univariate root-finding tool like `uniroot` in R.

A corresponding estimator for β is then simply given by

$$\hat{\beta}_n = \hat{m}/m(\hat{\xi}_n, 1) \quad (2.40)$$

so by construction `kMedMad` is equivariant in the sense of (1.5).

IF The implicit function of the two equations we have to solve in order to find the `kMedMAD` estimates is defined as follows:

$$G((\xi, \beta); (M, m)) = (G^{(1)}, G^{(2)})^\tau = \left(f_{m, \xi, \beta; k}(M), \beta \frac{2^\xi - 1}{\xi} - m \right)^\tau \quad (2.41)$$

By the implicit function theorem, the Jacobian in the Delta method is

$$D = - \left(\frac{\partial G}{\partial(\xi, \beta)} \right)^{-1} \frac{\partial G}{\partial(M, m)} \quad (2.42)$$

Then the influence function of `MedMAD` estimator is

$$\text{IF}(x; \text{MedMAD}, F) = D(\text{IF}(x; \text{kMAD}, F), \text{IF}(x; \text{Median}, F))^\tau \quad (2.43)$$

where the influence functions of median and MAD can be found in Rieder (1994, Chapter 1.5), and the one of `kMAD` is a simple generalization:

$$\text{IF}(x; m, F) = \left(\frac{1}{2} - \mathbb{I}(x \leq m) \right) / f(m) \quad (2.44)$$

$$\text{IF}(x; M, F) = \frac{\frac{1}{2} - \mathbb{I}(-M \leq x - m \leq kM)}{f(m+kM) - f(m-M)} + \frac{f(m+kM) - f(m-M)}{kf(m+kM) + f(m-M)} \frac{\mathbb{I}(x \leq m) - \frac{1}{2}}{f(m)} \quad (2.45)$$

while for the entries of D we note that

$$\begin{aligned} \frac{\partial G^{(1)}}{\partial \xi} &= -v \left(\frac{v^\xi - 1}{\xi^2} - \frac{1}{\xi} \log(v) \right) \Big|_{v=v_-}^{v_+}, & \frac{\partial G^{(1)}}{\partial \beta} &= \frac{v}{\xi \beta^2} (v^\xi - 1) \Big|_{v=v_-}^{v_+}, \\ \frac{\partial G^{(2)}}{\partial \xi} &= \frac{\beta}{\xi} \left(2^\xi \log(2) - \frac{2^\xi - 1}{\xi} \right), & \frac{\partial G^{(2)}}{\partial \beta} &= \frac{2^\xi - 1}{\xi}, \\ \frac{\partial G^{(1)}}{\partial M} &= \frac{kv_+^{\xi+1} + v_-^{\xi+1}}{\beta}, & \frac{\partial G^{(1)}}{\partial m} &= \frac{v_+^{\xi+1}}{\beta} \Big|_{v=v_-}^{v_+}, & \frac{\partial G^{(2)}}{\partial M} &= 0, & \frac{\partial G^{(2)}}{\partial m} &= -1 \end{aligned}$$

Again, we have equivariance,

$$\text{IF}_{(\xi, \beta)}(x; \text{kMedMAD}, F) = d_\beta \text{IF}_{(\xi, 1)}(x/\beta; \text{kMedMAD}, F) \quad (2.46)$$

ASV The asymptotic covariance of the `kMedMAD` estimator is

$$\text{asVar}(T) = D^T \Sigma D, \quad \Sigma = \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_{2,2} \end{pmatrix} \quad (2.47)$$

where with obvious generalizations, Σ may be read off from Serfling and Mazumder (2009) as the asymptotic covariance of median and kMAD:

$$\begin{aligned} a &= f(m-M) + f(m+kM), & b &= f(m-M) - f(m+kM), \\ c &= f(m-M) + kf(m+kM), & d &= b^2 + 4(1-a)bf(m), \end{aligned} \quad (2.48)$$

$$\begin{aligned} \sigma_{1,1} &= (4f(m))^{-2}, & \sigma_{2,2} &= f(m)^2(4c^2(f(m)^2 + d))^{-1} \\ \sigma_{1,2} &= \sigma_{2,1} = (4f(m)c)^{-1}(1 - 4F(m-M) + b/f(m)), \end{aligned} \quad (2.49)$$

ASB The IF of kMedMAD is bounded, so the asymptotic bias is finite.

FSBP We may again refer to R. and H. (2010b) where it is shown that

$$\varepsilon_n^* = \min\{\hat{N}'_n, \hat{N}''_n\}/n \quad (2.50)$$

for

$$\hat{N}'_n = \#\{X_i | \hat{m} < X_i \leq (k+1)\hat{m}\}, \quad (2.51)$$

$$\hat{N}''_n = \lceil n/2 \rceil - \#\{X_i | (1 - \check{q}_k)\hat{m} \leq X_i \leq (k\check{q}_k + 1)\hat{m}\} \quad (2.52)$$

Hence, by the usual LLN arguments,

$$\bar{\varepsilon}^* = \min\left(F_\theta((k+1)m) - \frac{1}{2}, F_\theta((k\check{q}_k+1)m) - F_\theta((1-\check{q}_k)m) - \frac{1}{2}\right) \quad (2.53)$$

For $\xi = 0.7$, the EFSBP is given by the first alternative if $k < 3.23$ and by the second one otherwise. On first glance, this would make for a “definition breakdown”, but if, e.g. we move the observations to $(k+1)\hat{m} - 0$, we obtain as estimator $\hat{\beta} = 0 + 0$ and $\hat{\xi} = \infty$, hence a breakdown in the original sense.

As to the choice of k , it turns out that a value of $k = 10$ gives reasonable values of ABP, asVar, asBias for a wide range of parameters ξ , as documented in Table 1. In the sequel this will be our reference value for k ; as to EFSBP, for $n = 40, 100, 1000$ we obtain $\bar{\varepsilon}_n^* = 29.16\%, 30.28\%, 30.94\%$, respectively (R. and H., 2010b, Table 2).

ξ	GES	GES ^{opt}	asVar	asVar ^{opt}	asMSE	asMSE ^{opt}	ABP	ABP ^{opt}
0.01	4.09	2.71	12.08	3.04	16.26	7.58	0.249	0.322
0.10	3.83	2.84	10.90	3.41	14.58	8.39	0.259	0.325
0.70	4.38	3.66	12.80	6.29	17.60	14.13	0.310	0.342
1.50	5.85	4.82	19.50	11.25	28.06	24.03	0.355	0.358
4.00	10.58	8.42	52.90	35.00	80.90	56.86	0.221	0.379

Table 1 Robustness properties of kMedMAD for $k = 10$ and several shape parameters compared to corresponding optimal values, i.e.; MBRE (GES), MLE (asVar), OMSE (asMSE), kMedMAD(k^{ABP}), $k^{\text{ABP}} = \text{argmax}_k \text{ABP}(\text{kMedMAD}(k))$ (ABP)

The results when optimizing kMedMAD in k w.r.t. the different robustness criteria for $\xi = 0.7$ can be looked up in R. and H. (2010a, Table 5).

Remark 2.5 Admittedly, for given k , eventually in n , $E_{(\xi, \beta)}[\varepsilon_n^*(\text{kMedMAD}(k))]$ is decreasing in ξ s.t. $\lim_{\xi \rightarrow \infty} E_{(\xi, \beta)}[\varepsilon_n^*(\text{kMedMAD}(k))] = 0$. At the same time, eventually in n , $\xi \mapsto E_{(\xi, \beta)}[\varepsilon_n^*(\text{PE}^*)]$ is increasing with $\lim_{\xi \rightarrow \infty} E_{(\xi, \beta)}[\varepsilon_n^*(\text{PE}^*)] = 1/4$. In particular, for $k = 10$, for $\xi \geq 4.964$, PE^* has a better EFSBP / ABP, in this case $\bar{\varepsilon}^*(\text{PE}^*) \geq 19.0\%$. But, eventually in n , the EFSBP of kMedMAD for the ABP-optimal $k^{\text{ABP}} = k^{\text{ABP}}(\xi)$ never drops below 32.1% for $\xi \in (0, 10]$ and below 25% for $\xi \in (0, 437]$, and achieves 39.9% for $\xi = 7.20$.

Hybrid Estimator Still, for small sample sizes we encounter failures to solve the corresponding equations for kMedMAD for $k = 10$ —8% for $n = 40$ and 2.3% for $n = 100$, compare Table 4 and R. and H. (2010a, Table 9). To lower this failure rate also in these cases, a hybrid estimator Hybr is used, that by default returns kMedMAD for $k = 10$, and by failure—tries out several values for k in a loop and returns the first estimator not failing: We start at $k = 3.23$ (producing maximal ABP), and then at each iteration multiply k by 3, and try out at most 20 k -values. This leads to failure rates of 2.3% for $n = 40$ and 0.0% for $n = 100$. Asymptotically, Hybr coincides with kMedMAD, $k = 10$.

2.7 Most bias-robust Estimator: MBRE

Minimizing the maximal bias on convex contamination neighborhoods, we obtain the MBRE estimator; in the terminology of Hampel et al. (1986) this is the *most B-robust* estimator. In our smooth situation, MBRE can also be obtained as a limit within the class of OBRE estimators, letting bias bound b tend to its minimum, the minimax bias ω_c^{\min} (see below).

Note however that contrary to Dupuis (1998), Dupuis and Field (1998) we use norm n_β from (1.7) to achieve the discussed invariance.

Its optimality is determined solely by its IF $\bar{\psi}$, the determining equations of which are given below. To this optimal IF, we have to find an ALE with $\bar{\psi}$ as influence function. This may be achieved in several ways (see Rieder (1994, chap. 6); in the literature most often M-estimators are used; we use a one-step construction, i.e. to a suitably consistent starting estimator $\theta_n^{(0)}$ (Hybr in our case), the corresponding ALE is defined as

$$\text{MBRE} = \theta_n^{(0)} + \frac{1}{n} \sum_{i=1}^n \bar{\psi}_{\theta_n^{(0)}}(X_i) \quad (2.54)$$

The IF minimizing asBias among all ALEs may be read off from Rieder (1994, Thm. 5.5.1(b)), its gross error sensitivity is given by

$$\omega_c^{\min} = \max \left\{ \text{tr} d_\beta^{-1} A d_\beta^{-1} / \text{En}_\beta(A\Lambda - a), \quad a \in \mathbb{R}^2, 0 \neq A \in \mathbb{R}^{2 \times 2} \right\} \quad (2.55)$$

while the optimal IF $\bar{\psi}$ is given by

$$\bar{\psi} = \omega_c^{\min}(A\Lambda - a) / n_\beta(A\Lambda - a) \quad (2.56)$$

where the event $\{A\Lambda - a = 0\}$ carries probability 0. Apparently, (2.56) only determines expression $A\Lambda - a$ up to a positive scalar multiple. For the values below, we have standardized this expression such that $A_{1,1} = 1$. There are no closed form expressions for A , a , and ω_c^{\min} , though. Corresponding algorithms to determine A , a , and ω_c^{\min} are implemented to R within the R0ptEst package Kohl and R. (2009) available on CRAN.

Remark 2.6 Although algorithms are implemented for general L_2 -differentiable models in R0ptEst, particular algorithms and techniques are needed for the computation of the expectations under GPD.

In our model, we obtain

$$A_{\text{MBRE}} = \begin{pmatrix} 1.00 & -0.18 \\ -0.18 & 0.22 \end{pmatrix}, \quad a_{\text{MBRE}} = (-0.18, 0.00), \quad \omega_c^{\min} = 3.67 \quad (2.57)$$

The use of norm n_β enforces (asympt.) in-/equivariance,

$$\bar{\psi}_{(\xi,\beta)}(x) = d_\beta \bar{\psi}_{(\xi,1)}(x/\beta) \quad (2.58)$$

or, suppressing subscript MBRE , with

$$Y_{(\xi,\beta)} = A_{(\xi,\beta)} \Lambda_{(\xi,\beta)}(x/\beta) - a_{(\xi,\beta)} \quad (2.59)$$

$$\begin{aligned} A_{(\xi,\beta)} &= d_\beta A_{(\xi,1)} d_\beta, & a_{(\xi,\beta)} &= d_\beta a_{(\xi,1)}, \\ n_\beta(Y_{(\xi,\beta)}) &= n_1(Y_{(\xi,1)}), & \omega_c^{\min}(\xi, \beta) &= \omega_c^{\min}(\xi, 1) \end{aligned} \quad (2.60)$$

2.8 Estimator minimizing maximal MSE: OMSE

For an estimator minimizing maximal MSE on neighborhoods (OMSE), we proceed similarly as for the MBRE: We determine the IF $\hat{\psi}$ of the corresponding optimal procedure and then use a one-step construction (with Hybr as starting estimator) to define an ALE with this IF as

$$\text{OMSE} = \theta_n^{(0)} + \frac{1}{n} \sum_{i=1}^n \hat{\psi}_{\theta_n^{(0)}}(X_i) \quad (2.61)$$

In the general L_2 differentiable setting, the form of $\hat{\psi}$ may be read off from Rieder (1994, Thm. 5.5.7):

$$\hat{\psi} = Y \min\{1, b/n_\beta(Y)\}, \quad Y = A\Lambda - a \quad (2.62)$$

where $A \in \mathbb{R}^{2 \times 2}$ and $a \in \mathbb{R}^2$ are such that $\hat{\psi}$ is an IF, i.e.; (1.13) holds, and b is such that

$$r^2 b = E(|Y| - b)_+ \quad (2.63)$$

Again, there are no closed form expressions for A , a , and b , but corresponding algorithms to determine A , a , and b are implemented to R within the R0ptEst package available on CRAN. In our model, we obtain

$$A_{\text{OMSE}} = \begin{pmatrix} 10.26 & -2.89 \\ -2.89 & 3.87 \end{pmatrix}, \quad a_{\text{OMSE}} = (-1.08, 0.12), \quad b_{\text{OMSE}} = 4.40 \quad (2.64)$$

Again, the use of norm n_β enforces (asympt.) in-/equivariance, i.e.; (2.58) holds mutatis mutandis, or again, (without the expression ω_c^{\min} and after suppressing OMSE), corresponding equations (2.59) and (2.60) together with

$$b_{(\xi,\beta)} = b_{(\xi,1)} \quad (2.65)$$

Remark 2.7 (a) OMSE also solves the ‘‘Lemma 5 problem’’ for bias bound its own GES (Rieder, 1994, Thm. 5.5.7), hence it is a particular OBRE in the terminology of Dupuis (1998), Dupuis and Field (1998). These authors, though, do not pursue the goal to find the MSE-optimal bias bound, and so our OMSE will in general beat their OBRE (w.r.t. MSE at our radius r). On the other hand, for given bias bound b , (2.63) also gives a radius $r(b)$ a given OBRE is MSE-optimal for; in this sense, bias bound b and radius r are equivalent parametrizations of the degree of robustness required for the solution.

(b) Passing to another risk does not in general invalidate our optimality (R. and Rieder, 2004): Whenever the asymptotic risk is representable as $G(\text{tr asVar}, |\text{asBias}|)$ for some convex function G isotone in both arguments, the optimal IF is again in the class of OBRE estimators—with possibly another bias weight. In addition, the radius-minimax procedure for MSE, i.e.; OMSE for $r = 0.486$ (Rem. 1.3) is simultaneously optimal for all homogenous risks according to Thm. 6.1 in the cited reference.

Computational Aspects Due to the lack of invariance in ξ , solving for equations (2.62) and (2.63) can be quite slow: for any new found starting estimate $\theta_n^{(0)}$ the solution has to be computed anew. Of course, we can reduce the problem by one dimension due to scale invariance, i.e.; we only would need to know the influence curves for ‘‘all’’ values $\xi > 0$. To speed up computation, especially for our simulation study, we therefore have used the following approximative approach, already realized in M. Kohl’s R package RobLox for the Gaussian one-dimensional location and scale model³, Kohl (2009):

Algorithm 2.8 For a grid of size M values of ξ , giving parameter values $\theta_i = (\xi_i, 1)$ and to given radius $r = 0.5$, we offline determine the optimal IF’s $\hat{\psi}_{\theta_i}$, solving equations (2.62) and (2.63) for each θ_i and suitably store the respective Lagrange multipliers A , a , and b , denoted by A_i , a_i , b_i . In the actual evaluation of OMSE at a given data set, for given starting estimate $\theta_n^{(0)}$, we reduce the problem by invariance and pass over to parameter value $\theta' = (\xi_n^{(0)}, 1)$. For this value, we find values A^\sharp , a^\sharp , and b^\sharp by simple inter-/extrapolation for the stored grid values A_i , a_i , b_i . This gives us $Y^\sharp = A^\sharp \Lambda_{\theta'} - a^\sharp$, and $w^\sharp = \min(1, b^\sharp/n_\beta(Y^\sharp))$. So far, $Y^\sharp w^\sharp$ would not satisfy (1.13) at θ' ; thus, similarly to Rieder (1994, Rem. 5.5.2), we generate an approximating IF ψ^\sharp by defining

$$z^\sharp = E_{\theta'}[\Lambda_{\theta'} w^\sharp] / E_{\theta'}[w^\sharp], \quad A^\sharp = \{E_{\theta'}[(\Lambda_{\theta'} - z^\sharp)(\Lambda_{\theta'} - z^\sharp)^\tau w^\sharp]\}^{-1}, \quad (2.66)$$

$a^\sharp = A^\sharp z^\sharp$, and $Y^\sharp = A^\sharp \Lambda_{\theta'} - a^\sharp$, and set $\psi^\sharp = \psi^\sharp w^\sharp$. By construction $E_{\theta'} \psi^\sharp = 0$ and $E_{\theta'} \psi^\sharp \Lambda_{\theta'}^\tau = \mathbb{I}_2$, so ψ^\sharp is indeed an IF at θ' .

Remark 2.9 ψ^\sharp produced in this way in general does not solve (2.62) and (2.63), i.e. $A^\sharp \neq A^\sharp$, $a^\sharp \neq a^\sharp$, nor holds equality in (2.63), but if the grid is dense enough, due to the smoothness of our model, we will have approximate equality in all these equations. For this smoothness (R. and H., 2010a, Figure 2). We have checked the accuracy in terms of efficiency loss w.r.t. the actual optimal IF in terms of relative asMSE: At the true parameter $\xi = 1$, we achieve 99.3% efficiency for OMSE and 99.0% for MBRE, while at $\xi = 0.1$, $\xi = 1.3$ we never drop below 99% efficiency.

The speed gain obtainable by Algorithm 2.8 is by a factor of roughly 125, and for larger n can be increased by yet another factor 10 if we may skip the re-centering/standardization and instead return $Y^\sharp w^\sharp$. We apply Algorithm 2.8 for both MBRE and OMSE.

³ Due to the affine equivariance of MBRE, OBRE, OMSE in the location and scale setting, interpolation in package RobLox is done only for varying radius r .

3 Synopsis of the Theoretical Properties

In a condensed form, in Table 2, we summarize our findings so far, evaluating criteria FSBP (where possible), asBias = r GES, trasVar, and asMSE (at $r = 0.5$). To give non-degenerate limits (in the shrinking neighborhood setting) and to be able to compare the results for different sample sizes n , these figures are standardized by the n (respectively \sqrt{n} for the bias). We also determine efficiencies eff.id, eff.re, and eff.ru. For FSBP of MLE, SMLE, we evaluate terms at sample size $n = 1000$, so $r' = 0.7$ entails $\alpha_n = 2.2\%$. Finally, we document the ranges of least favorable x -values $x_{l.f.}$, at which the considered IFs take their maximum in n_β -norm. Infinitesimally, these are the most vulnerable points of the resp. estimators, as contamination placing mass therein will render bias maximal. In all situations where $\infty \in x_{l.f.}$, 10^{10} will suffice to produce maximal bias in the displayed accuracy. On the other hand, Pickands estimator PE, as well as MMed are most harmfully contaminated by placing extra mass at smallish values of, say, about $x = 1.5$ (for $\beta = 1$).

The results for SMLE have to be read with care: asBias and asMSE do not account for the bias B_n already present in the ideal model, but only for the extra bias induced by contamination. Lemma 2.1 entails that B_n is of exact unstandardized order $O(\log(n)/\sqrt{n})$, hence consequently, asBias and asMSE should both be ∞ , and the efficiencies in ideal and contaminated situation would both be 0. For $n = 1000$, though, asBias and asMSE are finite: According to Lemma 2.1, $B_{1000} \approx 0.17$ (unstandardized), resp., 5.38, when multiplied by \sqrt{n} , while the entry of 3.75 in Table 2 is just GES and is at large due to an underestimation of ξ by 0.17.

As noted, MLE achieves smallest asVar, hence cannot be beaten in the ideal model, but at the price of a minimal FSBP and an infinite GES, so at any sample one large observation size suffices to render MSE arbitrarily large. kMedMAD gives very acceptable results in both asMSE and (E)FSBP; contrary to MDE, MLE, SMLE, MBRE, and OMSE it does not rely on a starting estimator though, as we only have to find zeros by univariate algorithms in canonically given search intervals.

The best breakdown behavior so far has been achieved by Hybr, with $\varepsilon^* \approx 1/3$ for a reasonable range of ξ -values. MDE shares an excellent reliability with Hybr, but contrary to the former needs a reliable starting value for the optimization. As to computation, it is quite fast though.

MBRE and OMSE are constructed as one-step estimators, so inherit the FSBP of the starting estimator (Hybr), while at the same time MBRE achieves lowest GES (unstandardized by n of order 0.1 at $n = 1000$), and OMSE is best according to asMSE; admittedly, though, MDE comes quite close in both efficiency and FSBP.

Considering unknown contamination radius and least favorable efficiency eff.ru, OMSE for $r = 0.5$ is best among the considered estimators and guarantees an efficiency of 0.68 over all radii. MDE, kMedMAD/Hybr, and MBRE also give acceptable least favorable efficiencies, never dropping considerably below 0.5, while all other estimators are less convincing.

In Figure 2, we display the influence curves (ICs) ψ_θ of the considered estimators. All of them are invariant so that $\psi_{(\xi,\beta)}(x) = d_\beta \psi_{(\xi,1)}(x/\beta)$.

estimator	asBias	trasVar	asMSE	eff.id	eff.re	eff.ru	$x_{l.f.}$	\mathcal{E}_{1000}^*
MLE	∞	6.29	∞	1.00	0.00	0.00	∞	0.00
PE	4.08	24.24	40.87	0.26	0.35	0.20	[0.89; 2.34]	0.06
MMed	2.62	17.45	24.32	0.36	0.58	0.32	[0.00; 0.34] \cup [0.90; 2.54]	0.25 [?]
kMedMAD	2.19	12.80	17.60	0.49	0.80	0.49	[0.54; 0.89] \cup [4.42; ∞)	0.31
SMLE	3.75	7.03	21.08	0.90	0.67	0.03	[20.67; ∞)	0.02
MDE	2.45	9.76	15.74	0.64	0.90	0.56	{0, ∞ }	0.35 [?]
MBRE	1.84	13.44	16.80	0.47	0.84	0.47	[0.00; ∞)	0.35*
OMSE	2.20	9.73	14.13	0.64	1.00	0.68	[0.00; 0.07] \cup [5.92; ∞)	0.35*

Table 2 Comparison of the asymptotic robustness properties of the estimators

*: inherited from starting estimator Hybr; ? : conjectured.

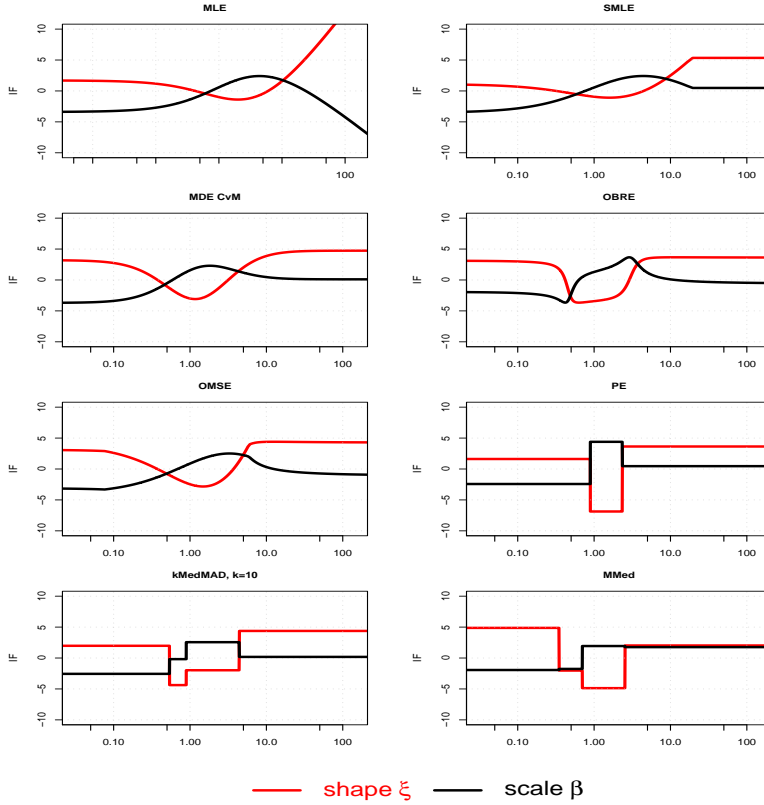


Fig. 2 Influence Functions

of MLE, SMLE (with $\approx 0.7 \cdot \sqrt{n}$ skipped value), MDE CvM, MBRE, OMSE, PE, MMed, kMedMAD estimators of the generalized Pareto distribution; mind the logarithmic scale of the x -axis

Intuitively, based on optimality within $L_2(P_\theta)$, in order to achieve high efficiency (in the ideal or contaminated situation), the IF should be as close as possible in L_2 -sense to the resp. optimal one. So, on first glance, it is astonishing, that kMedMAD achieves a reasonable efficiency in the contaminated situation, although its corresponding curves look quite different from the optimal ones of OMSE; but, of course, the difference occurs predominantly in regions of low F_θ -probability.

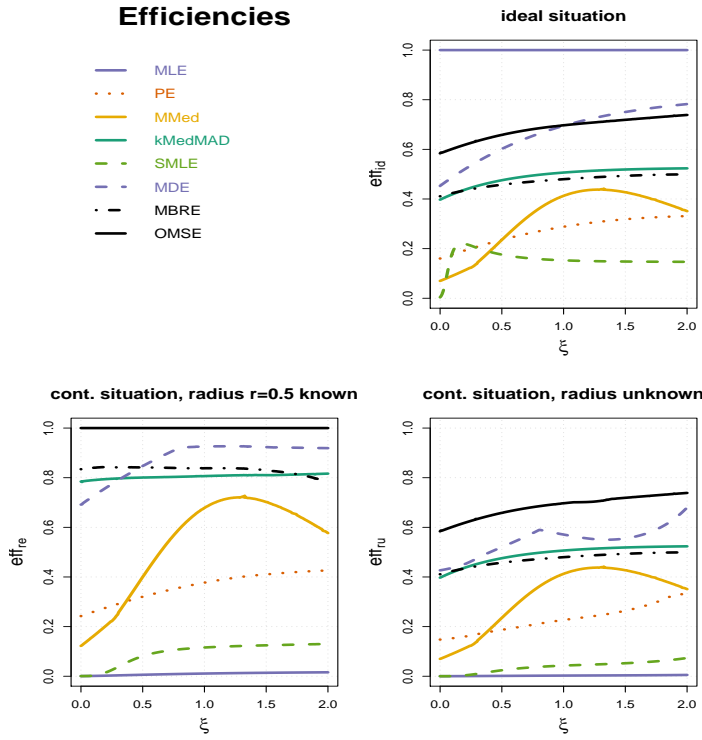


Fig. 3 Efficiencies for varying shape of MLE, SMLE (with $\approx 0.7 \cdot \sqrt{n}$ skipped value), MDE CvM, MBRE, OMSE, PE, MMed, kMedMAD estimators for scale $\beta = 1$ and varying shape ξ .

$\xi = 0.7$ is **typical**: Concerning the obtainable efficiencies, i.e. the conclusions we just have drawn as to the ranking of the procedures remain valid for other parameter values, as visible in Figure 3. Note that due to the scale invariance we do not need to consider $\beta \neq 1$. From this figure we may in particular read off the minimal value for the efficiencies as extracted in Table 3.

estimator	MLE	PE	MMed	kMedMAD	SMLE	MDE	MBRE	OMSE
$\min_{\xi} \text{eff.id}$	1.00	0.16	0.07	0.40	0.00	0.45	0.41	0.58
$\min_{\xi} \text{eff.re}$	0.00	0.24	0.12	0.78	0.00	0.69	0.78	1.00
$\min_{\xi} \text{eff.ru}$	0.00	0.15	0.07	0.40	0.00	0.43	0.41	0.58

Table 3 Minimal efficiencies for ξ varying in $[0, 2]$ in the ideal model and for contamination of known and unknown radius

4 Simulation Study

4.1 Setup

For sample size $n = 40$, we simulate data from both the ideal GPD with parameter values $\mu = 0$, $\xi = 0.7$, $\beta = 1$. Additional tables and plots for $n = 100, 1000$ can be found in R. and H. (2010a). We evaluate the estimators from the previous section at $M = 10000$ runs in the respective situation (ideal/contaminated and sample size $n = 40$).

The contaminated data stems from the (shrinking) Gross Error Model (1.9), (1.10) with $r = 0.5$. For $n = 40$, this amounts an actual contamination rate of $r_{40} = 7.9\%$. As contaminating data distribution, we use $G_{n,i} = \text{Dirac}(10^{10})$, except for estimators PE and MMed, where we use $G'_{n,i} = \text{unif}(1.42, 1.59)$ in accordance with $x_{1.f.}$ from Table 2. For MMed and kMedMAD, it turns out that, for maximal MSE we should use $G_{n,i}$ while $G'_{n,i}$ produces higher failure rates, so that in these two cases, for all entries except for the failure rate, we use $G_{n,i}$, and for column “NA” we use $G'_{n,i}$.

4.2 Results

Results are summarized in Tables 4. Values for n_β (Bias), tr Var , and for MSE (standardized by $\sqrt{40}$ and 40, respectively) all come with corresponding CLT-based 95%-confidence intervals. Column “NA” gives the failure rate in the computation in percent; basically, these are failures of MMed or kMedMAD to find a zero, which due to the use of Hybr as initialization is then propagated to MLE, SMLE, MDE, MBRE, and OMSE. Column “time” gives the aggregated computation time in seconds on a recent dual core processor for the 10000 evaluations of the estimator for ideal and contaminated situation. For MLE, SMLE, MDE, MBRE, and OMSE we do not include the time for evaluating the starting estimator (Hybr) but only write down the values for the evaluations given the respective starting estimate. The row with the respective best estimator is printed in bold face.

The simulation study confirms our findings of Section 3; figures are—at large—close to the ones of Table 2. This holds in particular for the ideal situation, and for the efficiencies, where in the latter case we obtain reasonable approximations already for $n = 100$ (R. and H., 2010a, Tables 8,9)—at the exception of SMLE and the PE-variants.

Essentially, the ranking given by asymptotics is valid already at sample size 40—as predicted by asymptotic theory, OMSE in its interpolated and IF-corrected variant at significance 95% is the best considered estimator as to MSE, although, especially for small sample sizes, MDE, MBRE, and Hybr come quite close as to efficiency in the contaminated situation.

Using Hybr as starting estimator, the number of failures can be kept low already at $n = 40$ —less than 1% in the ideal model and about 3% under contamination. This is not true for MMed and kMedMAD, which suffer from up to 33% failure rate at this n under contamination. So Hybr is a real improvement.

ideal situation:								
estimator	n_{β} (Bias)	tr Var	MSE	eff	NA	time		
MLE	0.55 ± 0.05	7.41 ± 0.21	7.72 ± 0.21	1.00	3.60	113		
PE	0.85 ± 0.27	19.30 ± 1.54	20.01 ± 1.67	0.39	0.00	13		
MMed	8.91 ± 1.98	1.02e5 $\pm 2.42e3$	1.02e5 $\pm 2.46e3$	0.00	10.44	168		
kMedMAD	0.47 ± 0.07	11.55 ± 0.30	11.78 ± 0.29	0.66	8.08	197		
Hybr	0.71 ± 0.07	11.96 ± 0.31	12.46 ± 0.30	0.62	0.79	223		
SMLE	4.70 ± 0.06	9.49 ± 0.30	31.62 ± 0.47	0.24	0.79	75		
MDE	0.40 ± 0.06	10.56 ± 0.27	10.72 ± 0.25	0.72	0.79	384		
MBRE	0.49 ± 0.08	15.68 ± 0.46	15.92 ± 0.44	0.48	0.79	302		
OMSE	0.26 ± 0.06	9.62 ± 0.23	9.68 ± 0.22	0.80	0.79	600		

contaminated situation:								
estimator	n_{β} (Bias)	tr Var	MSE	eff	NA			
MLE	394.12 ± 22.92	1.37e7 $\pm 1.20e6$	1.52e7 $\pm 1.37e6$	0.00	3.61			
PE	2.32 ± 0.49	62.25 ± 67.90	67.64 ± 69.35	0.39	0.00			
MMed	5.13 ± 1.17	3.56e3 $\pm 1.44e3$	3.59e3 $\pm 1.45e3$	0.01	23.11			
kMedMAD	2.32 ± 0.09	18.82 ± 0.49	24.21 ± 0.67	0.91	19.10			
Hybr	2.23 ± 0.09	19.23 ± 0.50	24.21 ± 0.67	0.91	3.03			
SMLE	7.44 ± 3.10	2.51e5 $\pm 1.52e5$	2.52e5 $\pm 1.52e5$	0.00	3.61			
MDE	2.64 ± 0.08	16.19 ± 0.43	23.15 ± 0.59	0.95	3.61			
MBRE	1.77 ± 0.09	20.06 ± 0.59	23.19 ± 0.63	0.95	3.03			
OMSE	2.75 ± 0.07	14.39 ± 0.42	21.93 ± 0.61	1.00	3.03			

Table 4 Comparison of the empirical robustness properties of the estimators at $n = 40$

The results for sample size 40 are illustrated in boxplots in Figures 4(a) and 4(b), respectively. In Figure 4(a), the underestimation of shape parameter ξ by SMLE in the ideal situation stands out; all other estimators in the ideal model are bias-free at large, while PE is somewhat less precise; under contamination (Figure 4(b)), all estimators are affected, producing bias, most prominently in coordinate ξ . As expected, this effect is most pronounced for MLE which is completely driven away, while the other estimators, at least in their medians stay near the true parameter value.

5 Conclusion

We have compared MLE, SMLE, MDE CvM, PE, MMed, kMedMAD, and the optimally robust MBRE and OMSE as estimators for scale and shape parameters ξ and β of the GPD on ideal and contaminated data in terms of local and global robustness properties.

Asymptotic theory and empirical simulations show that Hybr, kMedMAD, MDE, MBRE, and OMSE estimators can withstand relatively high outliers rates as expressed by $a(n)$ (E)FSBP of roughly $1/3$. SMLE in the variant without bias correction as used in this paper, but with shrinking skipping rate, and MLE have minimal FSBP of $1/n$, hence should be avoided.

High failure rates for MMed and kMedMAD for small n , and under contamination limit their usability considerably, while Hybr works reliably.

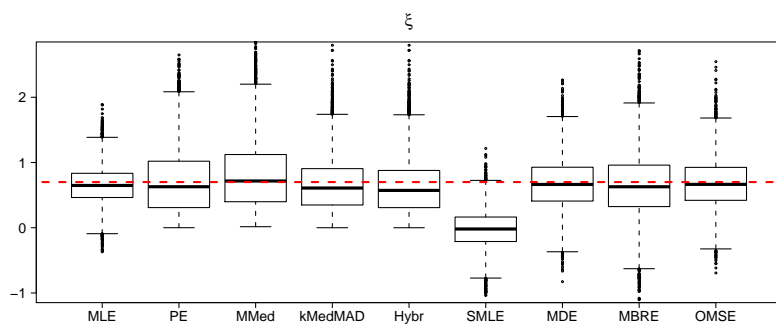
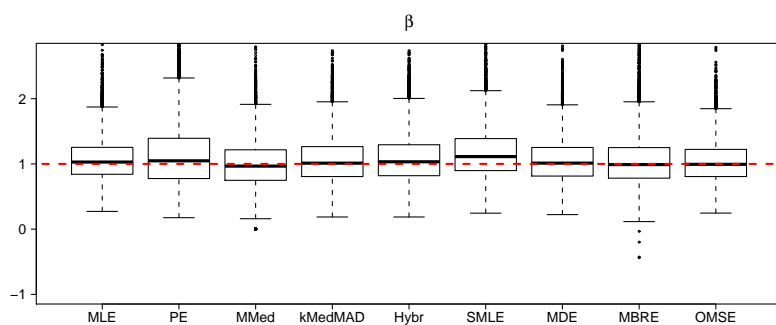
Looking at the influence functions, we see that, except for MLE, all estimators have bounded IFs, so finite GES. As visible in Figure 3, the estimators do differ

though in how they use the information present in an observation. This is reflected in asymptotic, as well as (simulated) finite sample risks: Overall, we can recommend OMSE with Hybr as initialization; it has achieved best risk in the simulations, may be computed fast, is efficient (100%) for contamination of known radius and, for $\xi \in [0, 2]$, never drops below 58% efficiency in the ideal model and for contamination of unknown radius (see Table 3). MBRE, and MDE come close to OMSE with minimal efficiencies $\text{eff.id} = \text{eff.ru} = 41\%$, $\text{eff.re} = 78\%$ (MBRE) and $\text{eff.id} = 45\%$, $\text{eff.re} = 69\%$, $\text{eff.ru} = 43\%$ (MDE). Among the potential starting estimators, clearly kMedMAD in its variant Hybr stands out and comes closest to the aforementioned group— $\text{eff.id} = \text{eff.ru} = 40\%$, $\text{eff.re} = 78\%$. PE is also robust, but not really advisably due to its low breakdown point and non-convincing efficiencies; the only reason for using PE is its ease of computation, which should not be so decisive, though. Even worse is the popular SMLE without bias correction, which does provide some, but much too little protection against outliers. Worst, of course, as to all robustness aspects is MLE.

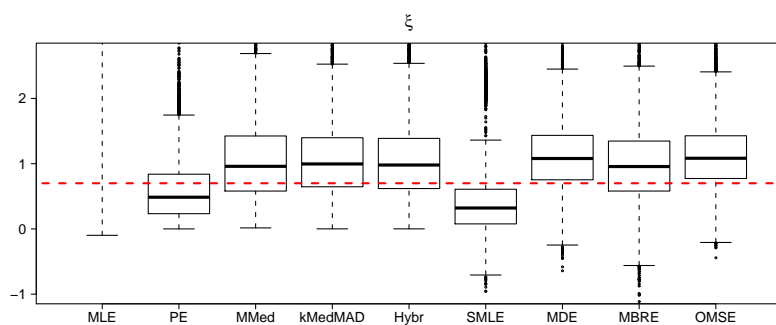
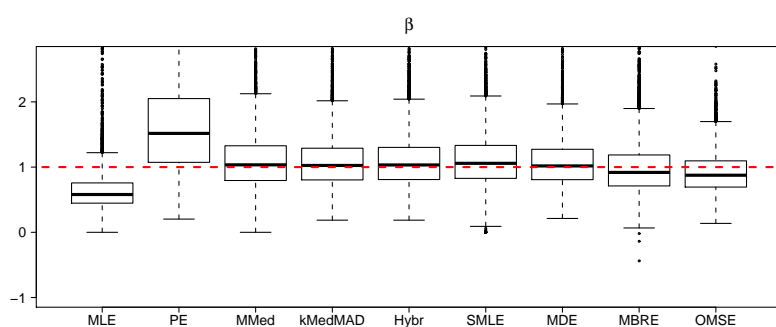
References

- Basel Committee on Banking Supervision: International Convergence of Capital Measurement and Capital Standards: A Revised Framework. <http://www.bis.org/publ/bcbs128.pdf> (2006)
- Balkema, A., de Haan, L.: Residual life time at great age. *Annals of Probability* 2, 792–804 (1974)
- Brazauskas, V., Serfling, R.: Robust Estimation of Tail Parameters for Two-Parameter Pareto and Exponential Models via Generalized Quantile Statistics. *Extremes* 3(3), 231–249 (2000)
- Bickel, P.J.: Quelques aspects de la statistique robuste. In P.L. Hennequin (ed.) *Ecole d'Été de Probabilités de Saint Flour IX 1979*, Lecture Notes in Mathematics #876, pp. 1–72, Springer (1981)
- Castillo, E., Hadi, A. S.: Fitting the Generalized Pareto Distribution to Data. *JASA* 92(440), 1609–1620 (1997)
- Cope, E. W., Mignola, G., Antonini, G., Ugocioni, R.: Challenges and pitfalls in measuring operational risk from loss data. *Journal of Operational Risk* 4(4) (2009)
- Donoho, D. L. and Huber, P. J. The notion of breakdown point. In Bickel, P. J., Doksum K., and Hodges, J. L. Jr. (eds) *A Festschrift for Erich L. Lehmann*, pp. 157–184, Wadsworth, Belmont, CA. (1983)
- Donoho, D. L., Liu, R. C.: The “Automatic” Robustness of Minimum Distance Functionals. *Annals of Statistics* 16(2), 552–586 (1988)
- Dupuis, D. J.: Exceedances over high thresholds: A guide to threshold selection. *Extremes* 1(3), 251–261 (1998)
- Dupuis, D. J., Field, C. A.: Robust estimation of extremes. *Can. J. Stat.* 26, 199–216 (1998)
- Dupuis, D. J., Morgenthaler, S.: Robust weighted likelihood estimators with an application to bivariate extreme value problems. *Can. J. Stat.* 30(1), 17–36 (2002)
- Fernholz, L. T.: von Mises Calculus for Statistical Functionals. *Lecture Notes in Statistics* #19. Springer (1979)
- Field, C., Smith, B.: Robust Estimation—A Weighted Maximum Likelihood Estimation. *International Review* 62(3), 405–424 (1994)
- Hadi, A.S., Luceño, A.: Maximum trimmed likelihood estimators: a unified approach, examples, and algorithms. *CSDA* 25, 251–272 (1997)
- He, X., Fung, W. K.: Method of medians for life time data with Weibull models. *Statistics in Medicine* 18, 1993–2009 (1999)
- Hampel, F. R.: Contributions to the theory of robust estimation. Ph.D. Thesis, University of California, Berkeley (1968)
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., Stahel, W. A.: Robust statistics. The approach based on influence functions. Wiley (1986)

-
- Hosking R. J. M., Wallis T. J.: Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics* 29(3) (1987)
- Huber, P. J.: *Robust statistics*, Wiley (1981)
- Huber-Carol, C.: Étude asymptotique de tests robustes. PhD Thesis, ETH Zürich (1970)
- Juárez, S. F.: Robust and efficient estimation for the generalized Pareto distribution. Unpublished Ph.D. dissertation, Statistical Science Department, Southern Methodist University. <http://www.smu.edu/statistics/faculty/SergioDiss1.pdf> (2003)
- Juárez, S. F., Schucany, W. R.: Robust and Efficient Estimation for the Generalized Pareto Distribution. *Extremes* 7(3), 237–251 (2004)
- Kohl, M.: Numerical Contributions to the Asymptotic Theory of Robustness. PhD Thesis, Universität Bayreuth. <http://stamats.de/ThesisMKohl.pdf> (2005)
- Kohl, M.: RobLox: Optimally robust influence curves and estimators for location and scale. R Package available in version 0.7 on CRAN, <http://cran.r-project.org/> (2009)
- Kohl, M., Rieder H., Ruckdeschel, P.: Infinitesimally Robust Estimation in General Smoothly Parametrized Models. To appear in *Stat. Meth. & Appl.* (2010)
- Kohl, M., Ruckdeschel, P.: ROptEst: Optimally robust estimation. R Package available in version 0.7 on CRAN, <http://cran.r-project.org/> (2009)
- Müller, C. H., Neykov, N.: Breakdown points of trimmed likelihood estimators and related estimators in generalised linear models. *J. Statist. Plann. Inference* 116, 503–519 (2003)
- Neslehova, J., Chavez-Demoulin, V., Embrechts, P.: Infinite Mean models and the LDA for operational risk. *Journal of Operational Risk* 1(1), 3–25 (2006)
- Peng, L., Welsch, A. H.: Robust Estimation of the Generalized Pareto Distribution. *Extremes* 4(1), 53–65 (2001)
- PICKANDS, J.: Statistical Inference Using Extreme Order Statistics. *Annals of Statistics* 3(1), 119–131 (1975)
- R Development Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org> (2009)
- Ribatet, M.: POT: Generalized Pareto Distribution and Peaks Over Threshold. R package, version 1.1-0 <http://cran.r-project.org/> (2009)
- Rieder, H.: *Robust Asymptotical Statistics*. Springer (1994)
- Rieder, H.: A robust asymptotic testing model. *Annals of Statistics* 6, 1080–1094 (1978)
- Rieder, H., Kohl, M., Ruckdeschel, P.: The Cost of not Knowing the Radius. *Stat. Meth. & Appl.*, 17(1), 13–40 (2008)
- Ruckdeschel P., Horbenko, N.: Robustness Properties of Estimators in Generalized Pareto Models. Technical Report No. 182, Fraunhofer ITWM, Kaiserslautern (2010a)
- Ruckdeschel P., Horbenko, N.: Yet another breakdown point notion: EFSBP. Preprint. Submitted ArXiv No. (2010b)
- Ruckdeschel, P. and Rieder, H. Optimal influence curves for general loss functions. *Statistics and Decisions* 22: 201–223 (2004)
- Serfling, R., Mazumder, S.: Exponential Probability Inequality and Convergence Results for the Median Absolute Deviation and Its Modifications. *Statistics and Probability Letters* 79, 1767–1773 (2009)
- Smith, L. R.: Estimating tails of probability distributions. *Annals of Statistics* 15(3), 1174–1207 (1987)
- van der Vaart, A. W.: *Asymptotic statistics*. Cambridge Univ. Press, Cambridge (1998)
- Vandev, D. L., Neykov, N. M.: About regression estimators with high breakdown point. *Statistics* 32, 111–129 (1998)
- Witting, H.: *Mathematische Statistik I: Parametrische Verfahren bei festem Stichprobenumfang*. B.G. Teubner, Stuttgart (1985)



(a) no contamination, 40 sample size



(b) 7.9 % contamination, 40 sample size

Fig. 4 Boxplots for MLE, PE, MMed, kMedMAD, Hybr, SMLE (with $\approx 0.7 \cdot \sqrt{n}$ skipped values), MDE, MBRE, OMSE estimators for shape ξ and scale β of the generalized Pareto distribution on the ideal (above) and contaminated data (below), (a), (b), number of simulations: 10000; the red dashed line is the true parameter value.