Fraunhofer
ITWM

P. Ruckdeschel, N. Horbenko

Robustness Properties of Estimators in Generalized Pareto Models

Berichte des Fraunhofer ITWM, Nr. 182 (2010)

ISSN 1434-9973

Bericht 182 (2010)

Die Veröffentlichungen in der Berichtsreihe des Fraunhofer ITWM können bezogen werden über:

Fraunhofer-Institut für Techno- und
Wirtschaftsmathematik ITWM
Fraunhofer-Platz 1

67663 Kaiserslautern
Germany

Telefon: +49(0)631/31600-0
Telefax: +49(0)631/31600-1099
E-Mail: info@itwm.fraunhofer.de
Internet: www.itwm.fraunhofer.de

Vorwort

Das Tätigkeitsfeld des Fraunhofer-Instituts für Techno- und Wirtschaftsmathematik ITWM umfasst anwendungsnahe Grundlagenforschung, angewandte Forschung sowie Beratung und kundenspezifische Lösungen auf allen Gebieten, die für Techno- und Wirtschaftsmathematik bedeutsam sind.

In der Reihe »Berichte des Fraunhofer ITWM« soll die Arbeit des Instituts kontinuierlich einer interessierten Öffentlichkeit in Industrie, Wirtschaft und Wissenschaft vorgestellt werden. Durch die enge Verzahnung mit dem Fachbereich Mathematik der Universität Kaiserslautern sowie durch zahlreiche Kooperationen mit internationalen Institutionen und Hochschulen in den Bereichen Ausbildung und Forschung ist ein großes Potenzial für Forschungsberichte vorhanden. In die Berichtreihe werden sowohl hervorragende Diplom- und Projektarbeiten und Dissertationen als auch Forschungsberichte der Institutsmitarbeiter und Institutsgäste zu aktuellen Fragen der Techno- und Wirtschaftsmathematik aufgenommen.

Darüber hinaus bietet die Reihe ein Forum für die Berichterstattung über die zahlreichen Kooperationsprojekte des Instituts mit Partnern aus Industrie und Wirtschaft.

Berichterstattung heißt hier Dokumentation des Transfers aktueller Ergebnisse aus mathematischer Forschungs- und Entwicklungsarbeit in industrielle Anwendungen und Softwareprodukte – und umgekehrt, denn Probleme der Praxis generieren neue interessante mathematische Fragestellungen.

Prof. Dr. Dieter Prätzel-Wolters
Institutsleiter

Kaiserslautern, im Juni 2001

Fraunhofer
ITWM

Robustness Properties of Estimators in Generalized Pareto Models

Peter Ruckdeschel
peter.ruckdeschel@itwm.fraunhofer.de
Fraunhofer ITWM, Department of Financial Mathematics and
Department of Mathematic, Statistics Group
TU Kaiserslautern, POB 3049, D-67653 Kaiserslautern

Nataliya Horbenko
nataliya.horbenko@itwm.fraunhofer.de
Department of Mathematic, Statistics Group and
Fraunhofer ITWM, Department of Financial Mathematics
Fraunhofer-Platz 1, D-67663 Kaiserslautern

March 24, 2010

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, moment-based estimators, Cramér-von-Mises Minimum Distance estimators, and, as a special case of quantile-based estimators, Pickands Estimator as well as variants of the latter tuned for higher finite sample breakdown point (FSBP), and lower variance.

We further consider an estimator matching population median and median of absolute deviations to the empirical ones (MedMad); again, in order to improve its FSBP, we propose a variant using a suitable asymmetric Mad as constituent, and which may be tuned to achieve 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 FSBP, 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

Contents

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 which is motivated by the Pickands-Balkema-de Haan extreme value theorem (PBHT), compare 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**, by definition "the risk of direct or indirect loss resulting from inadequate or failed internal processes, people and systems or from external events". In quantifying this risk, usually the tail behavior of the underlying distribution as expressed by tail quantiles (e.g., VaR) or truncated moments (CVaR) is crucial. Estimating these population quantiles by their empirical counterparts apparently is drastically prone to outliers: For the 99.9% quantile as typically used in this context, Basel II (2006), for 5000 observations, five irreproducible, extraordinarily large observations suffice to render this procedure completely meaningless. This is where extreme value theory enters, suggesting to estimate these quantiles parameterically 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. For the GPD, this unboundedness is induced by the shape parameter which decides upon the tail behavior of the distribution. Thus, in our example, five outliers will still invalidate our estimation.

Robust Statistics in contrast offers procedures bounding the influence of single observations, so provides reliable inference in the presence of moderate deviations of the underlying model assumptions, respectively the mechanisms underlying the PBHT. Admittedly, this comes at the price of some efficiency loss in the ideal model, which in practice may also be a problem.

In this article, we pick up certain estimators proposed as robust alternatives to MLEs and moment-based estimators in the literature. We examine their behavior in ideal and non-ideal situations as to their robustness and efficiency properties; based on these properties we give some indications on how to improve these estimators, and introduce some new ones with even better robustness properties.

Literature Estimating the three-parameter GPD has been a challenging problem for statisticians for many years, and many approaches to fit the GPD to real data have been proposed.

The MLE for the GPD is very popular for practitioners, as justified by its asymptotic optimality in terms of efficiency at the smooth model in the sense of the asymptotic Cramér Rao bound (which is restricted to asymptotically linear estimators, though). This estimator has been studied in detail by Smith (1987). For finite sample sizes, this optimality may not yet hold: Hosking & Wallis (1987) already note that the MLE in this case turns out to be inefficient even for large sample sizes compared against moment-based estimators.

To stabilize this procedure, Cope et al. (2009) propose 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). It is defined as the solution of a "Lemma 5 problem" (alluding to Lemma 5 of Hampel (1968)), i.e.; to a given bound on the bias in the neighborhood (more specifically, a bound b on the gross error sensitivity GES as defined in (1.20)) of the influence function, minimize the trace of the variance (cf. (Hampel et al., 1986, 2.4 Thm. 1)).

Generalizing He and Fung (1997), Peng and Welsch (2001) propose a method of median estimator which is based on solving the implicit equations matching the population medians of the coordinates of the scores function to the data; it is shown that this estimator is related to but not identical to the MBRE estimator we introduce later; one might hope that, as in the Weibull setting of He and Fung (1997), the asymptotic breakdown point of this procedure would be 50%, but no such result is derived in the cited reference.

All methods so far involve solving implicit equations, hence depend on suitable initializations. This is not true for the Elementary Percentile Method (EPM) introduced by Castillo and Hadi (1997) which applies quantile-based estimators to produce $\sqrt{n}$-consistent estimators, and as special case gives Pickands estimator (PE), Pickands (1975). Compared to the other methods, EPM estimators also may be computed much faster.

The approach by Brazauskas & Serfling (2000) uses a different parametrization of the GPD, i.e., if observations $X_i \overset{\text{i.i.d.}}{\sim} \text{GPD}(\beta, \xi)$ in our notation, they instead consider observations $Y_i = X_i + \beta/\xi$ and parametrize their model by $\alpha = \xi^{-1}$ and $\sigma = \beta/\xi$. In their setting, $\mathcal{L}(\log(Y_i)) = \mathcal{L}(\log(\beta/\xi) + E/\xi)$,

$E \sim \mathrm{Exp}(1)$, 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.

For shape parameter $\xi \leq 0.5$, second moments exist, and then moment-based methods such as the Method of Moments and the Method of Probability Weighted Moments (MPWM), Hosking & Wallis (1987) can be applied, and, for finite sample size, in the ideal model, behave quite competitive.

Juárez and Schucany (2004), Juárez (2003) apply a Minimum Density Power Divergence (distance) Estimator (MDPDE). An additional tuning parameter allows for defining the distance between the empirical and theoretical distributions one has to minimize in order to find the estimates.

We do not consider MPWM and MDPDE estimators in this paper, though.

None of the mentioned approaches gives a cure-all procedure: Depending on the loss function and on how large the deviation from the ideal model may be, the ranking among the alternatives may vary.

Estimators considered in this paper (for actual definitions see section 2):
- the Maximum Likelihood Estimator (MLE)
- the Skipped Maximum Likelihood Estimator (SMLE)
- the classical (first and second) moment-based estimator (MME)
- the Cramér-von-Mises Minimum Distance estimator (MDE)
- Pickands Estimator (PE) as a special case of quantile-based estimators
- variants of PE to achieve lower variance (PicM), resp. maximal breakdown point (among PE-type estimators): (PE*)
- the Method-of-Median estimator of Peng and Welsch (2001) (MMed)
- an estimator based on median and median of absolute deviations (Mad), (MedMad)
- a variant of MedMad (kMedMad) based on a suitably asymmetric Mad to achieve a high breakdown point and, at the same time close-to-optimal MSE behavior on neighborhoods
- the optimally-bias-robust estimator minimizing the maximal bias (MBRE)
- the estimator minimizing the maximal MSE (OMSE)

All of these estimators are asymptotically linear, hence asymptotically normal.

We have selected the procedures for the following reasons: MLE, MBRE, OMSE are optimal in certain settings, so serve as benchmarks. Pickands-type estimators, MMed, MedMad, and kMedMad are candidates for (robust) initialization estimators. MME is an example for a procedure even less robust than MLE, and SMLE, MDE have already been used in our application, hence are competitors.

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

Structure of the paper In section 1.1, we outline the generalized Pareto distribution and, for the deviations from this model, we define contamination neighborhoods, known as Gross Error Model.

To cope with these model deviations, in section 1.2, we recall global (finite sample breakdown point) and local (influence function) robustness criteria for estimators, together with efficiency measures such as asymptotical bias, variance, and mean squared error (MSE). In this context, we introduce the new concept of an *expected finite sample breakdown point* in Definition 1.6. Subsequently, section 2 describes the properties of the above-mentioned estimators: We analytically calculate the influence functions and asymptotic measures for MLE, SMLE, MME, PE, PE*, PicM, MedMad, kMedMad, and MDE, and, numerically, for MMed, MBRE, and OMSE estimators.

As already noted, MLE, MME, MDE, MMed, and PE have already been studied and their influence functions, asymptotic variances determined by other authors. We hence only cite the corresponding expressions, correcting some errors in the references. In addition to the cited literature, we introduce a new variant of Pickands estimator, PicM, which achieves a good compromise of variance and robustness. Also we contribute the MedMad estimator and its variant kMedMad, both of which, to our knowledge, are novel. Finally, in the context of Pareto distributions MBRE and OMSE have not yet been compared to the cited estimators as to their asymptotic variances, and maximal MSEs.

The main contribution of this paper is a synopsis section 3 where in tables and graphics we summarize our findings at a reference parameter setting. 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) is given by its c.d.f. and density

$$F_\theta(x) = 1 - \left(1 + \xi \frac{x-\mu}{\beta}\right)^{-\frac{1}{\xi}}, \quad f_\theta(x) = \frac{1}{\beta}\left(1 + \xi \frac{x-\mu}{\beta}\right)^{-\frac{1}{\xi}-1},$$
$$x \geq \mu \; (\xi \geq 0), \quad \mu < x \leq \mu - \frac{\beta}{\xi} \; (\xi < 0) \tag{1.1}$$

for parameters μ (location), $\beta > 0$ (scale) and ξ (shape). Special cases of GPDs are the uniform ($\xi = -1$), the exponential ($\xi = 0$, $\mu = 0$), and Pareto ($\xi > 0$, $\beta = 1$) distributions. According to our application, we limit ourselves to the case $\xi > 0$ here. Note that for the quantile function of a GPD the following relation holds:

$$f(F^{-1}(\alpha)) = \beta^{-1}(1-\alpha)^{1+\xi} \tag{1.2}$$

GPD is a good candidate for modeling the distributional tails from the threshold point μ on as motivated by the Pickands-Balkema-de Haan extreme value theorem, compare Balkema and de Haan

(1974), Pickands (1975), which states that for distributions in the maximal attraction domain of the Fréchet distribution with parameter ξ, the (suitably standardized and centered) exceedances over a high threshold are asymptotically distributed according to a GPD with shape parameter ξ; we will not use this argument in this paper, though.

Unfortunately, estimating the location parameter μ induces non-smoothness into the model: It can be shown that the corresponding model including μ is not L_2-differentiable (compare (van der Vaart, 1998, last paragraph, p. 129) and Horbenko & Ruckdeschel (2010)) which can be understood heuristically, as we do not see observations smaller than μ, and hence, similar to estimating parameter θ in $\mathrm{uniform}(0, \theta)$, the minimal observation (with non-CLT-asymptotics) will be sufficient for estimating μ.

In applications, data for fitting the GPD is obtained in a two-step procedure: As the GPD is only used to fit the tail of the data, in a first step, the threshold μ is determined ("estimated")[1]. This is by no means trivial: According to theory, the threshold point should be set high enough to fit the tail of the distribution with GPD, but should also be low enough to leave us a sufficient amount of data beyond that threshold for the estimation of the other parameters, i.e. shape ξ and scale β. For given threshold μ then, in a second step, the reduced model (only in (ξ, β)) is fitted.
In practice the first step amounts to looking for "flat" regions in a corresponding threshold plot: a plot of the function $\mu \to \hat{\theta}(\mu)$ with $\hat{\theta}(\mu)$ being an estimator depending on the threshold μ (Baud et al., 2002).

For this article, we limit ourselves to the second step, assuming the location parameter μ to be known and equal to zero. For all graphics and numerical and simulational evaluations we use the reference parameter values $\beta = 1$ and $\xi = 0.7$.

After this reduction, the model is smooth, i.e. L_2-differentiable (compare (Witting, 1985, Satz 1.194), (van der Vaart, 1998, Definition (5.38))), as the density f_θ is differentiable in θ and the corresponding Fisher information is finite and continuous in θ, with L_2-derivative

$$\Lambda_\theta(z) = \left(\frac{1}{\xi^2}\log(1+\xi z) - \frac{\xi+1}{\xi}\frac{z}{1+\xi z}; -\frac{1}{\beta} + \frac{\xi+1}{\beta}\frac{z}{1+\xi z}\right)^\tau, \quad z = \frac{x-\mu}{\beta} \tag{1.3}$$

For integrations it turns out useful to introduce

$$v^{-\xi} = 1 + \xi z \tag{1.4}$$

and $\Lambda_\theta(z) =: \tilde{\Lambda}_\theta(v(z))$ defined as

$$\tilde{\Lambda}_\theta(v) = \left(-\frac{1}{\xi}\log(v) - \frac{\xi+1}{\xi^2}(1-v^\xi); -\frac{1}{\beta} + \frac{\xi+1}{\beta\xi}(1-v^\xi)\right)^\tau \tag{1.5}$$

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

[1] This is, according to S. Resnick the "black art" of extreme value statistics.

Using this quantile-type transformation, we easily obtain the Fisher information matrix $\mathcal{I}_\theta \in \mathbb{R}^{2\times 2}$ as

$$\mathcal{I}_\theta = \frac{1}{(2\xi+1)(\xi+1)} \begin{pmatrix} 2 & \beta^{-1} \\ \beta^{-1} & \beta^{-2}(\xi+1) \end{pmatrix} \tag{1.6}$$

We note that $\mathcal{I}$ is positive definite for our parameter range, hence the model is (locally) identifiable.

The reduced model enjoys a certain *invariance*: with an included scale component, it remains invariant under scale transformations $s_\beta(x) = \beta x$ of the observations. Using matrix

$$d_\beta = \operatorname{diag}(1, \beta) \tag{1.7}$$

this invariance is reflected by a corresponding notion of equivariance of estimators, i.e., an estimator S for $\theta = (\xi, \beta)$ is called *(scale)-equivariant* if

$$S(\beta x_1, \ldots, \beta x_n) = d_\beta S(x_1, \ldots, x_n) \tag{1.8}$$

In terms of the L_2 derivative, this invariance is reflected by

$$\Lambda_{(\xi,\beta)}(z) = d_\beta^{-1} \Lambda_{(\xi,1)}(z) \tag{1.9}$$

To preserve this invariance when determining the "length" of a parameter, Robust Statistics has used norms for the parameter space based on the Fisher information or on the respective asymptotic covariance matrix—see (Hampel et al., 1986, 4.2 Def.'s 3 and 4) giving so-called information—resp. self-standardized influence functions. Instead we propose a simpler invariant norm, based on d_β: For given parameter β, we use the weighted norm

$$n_\beta(x, y) = \|d_\beta^{-1}(x, y)\| = \sqrt{x^2 + y^2/\beta^2} \tag{1.10}$$

which also has the advantage that for large scale β the corresponding scale component of the estimator does not obtain an overly high weight.

Remark 1.1. For the shape parameter there is no obvious such invariance, except for the quantile transformation, of course, i.e., the transformation

$$g(\theta, \theta'; x) = F_{\theta'}^{-1} \circ F_\theta(x) = \left[(1 + \xi x/\beta)^{\xi'/\xi} - 1\right]\beta'/\xi' \tag{1.11}$$

transforming an F_θ-distributed observation X into an $F_{\theta'}$-distributed one. The only values of x which stay 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 with ideal distributions, in Robust Statistics, for some given size or radius $\varepsilon > 0$, one defines suitable distributional neighborhoods about this ideal model. In this paper, we limit ourselves to the *Gross Error Model*, i.e.; as neighborhoods, we use the set of all distributions F^{re} representable as:

$$F^{\mathrm{re}} = (1 - \varepsilon)F^{\mathrm{id}} + \varepsilon F^{\mathrm{di}} \tag{1.12}$$

where F^{id} is the underlying ideal distribution and F^{di} some arbitrary, unknown, and uncontrollable contaminating distribution. In the shrinking neighborhood approach as developed (a.o.) in Huber-Carol (1970), Rieder (1978), Bickel (1981), and Rieder (1994), in order to balance bias and variance (of different scaling otherwise) one lets the radius of these neighborhoods shrink with growing sample size n, i.e.,

$$\varepsilon = r_n = r/\sqrt{n} \tag{1.13}$$

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

In reality one rarely knows ε or r. Objective criteria for the choice of this radius (fixed or shrinking) for specifying a procedure in situations where one has no or only limited knowledge of the "true" radius are given in the Minimax method of Rieder et al. (2008). For our numerical and simulational evaluations, we use a starting radius $r = 0.5$.

Remark 1.2. Starting radius $r = 0.5$ actually almost is the minimax radius in the situation where we have no knowledge at all about the radius, which for our reference parameter $\theta = (\xi = 0.7, \beta = 1)$ would be 0.486, leading to a maximin efficiency of 0.683, i.e. using the corresponding radius minimax procedure, (with a clipping of $b = 4.436$) the performance of this procedure would never be worse than 1.464 times the maximal asymptotic mean squared error 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 $r = 0$), so very close to optimal.

1.2 Robustness

The robustness concepts used in this paper may be distinguished into local (measuring the influence of a single observation, for infinitesimally small deviations) and global ones (measuring the effect of massive deviations). The most important local robustness concept is the *influence function* (IF), while for the global concepts we recur to the *breakdown point*.

Influence Function Defining the estimator via a functional T evaluated at the empirical distribution, one can specify the infinitesimal influence of the individual observations on the estimator: The IF 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

$$\mathrm{IF}(x; T, F) = \lim_{\varepsilon \to 0} \frac{T(F_\varepsilon) - T(F)}{\varepsilon} \tag{1.14}$$

This definition however is too weak to support the chain rule which ought to be a minimal requirement for many applications. Thus, in fact stronger concepts like Hadamard or Bouligand derivatives are needed (for the use of the latter in this context, see, e.g. Christmann and Van Messem (2008)), and fortunately corresponding results can be read off from Fernholz (1979), (Rieder, 1994, chap. 1) for our estimators.

Using the (finite-dimensional) 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 cumulative distribution function (MDE), all settled in the cited references, while the results on one-step estimators of (Rieder, 1994, Chap. 6) suffice to show that MBRE and OMSE do have an influence curve.

According to Kohl et al. (2010), we would like to point out though, that the interpretation as the infinitesimal influence of a single observation for the estimator can also be obtained in a conceptionally simpler way, bypassing derivative notions: Assuming an L_2-differentiable model, one only checks that the estimator S_n has an expansion in the observations X_i as

$$S_n = \theta' + \frac{1}{n}\sum_{i=1}^{n} \psi_\theta(X_i) - \mathrm{E}_{\theta'}[\psi_\theta] + R_n, \qquad \sqrt{n}\,|R_n| \stackrel{n\to\infty}{\longrightarrow} 0 \quad P^n_{\theta'}\text{-stoch.} \tag{1.15}$$

where the influence function IF_θ of S_n is just $\psi_\theta - \mathrm{E}_\theta[\psi_\theta]$ for some function $\psi_\theta \in L_2(P_\theta)$, and (1.15) holds for all θ' s.t. $|\theta' - \theta| = \mathrm{O}(n^{-1/2})$. An estimator with (1.15) is called *asymptotically linear* or ALE.

We already note that all estimators considered in this paper are ALEs.

In Rieder (1994), contrary to other references, one imposes two side conditions for an $\mathrm{IF}\ \psi$: one works in the setup of L_2-differentiable models and requires that $\mathrm{E}\,\psi = 0$ and $\mathrm{E}\,\psi\Lambda^\tau = \mathbb{I}_k$; this may be motivated by the following lemma in the spirit of (Rieder, 1994, Lemma 4.2.18):

Lemma 1.3. *For $\Theta \subset \mathbb{R}^k$ an open parameter domain, let $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ a parametric model. Assume $\mathcal{P}$ is L_2-differentiable in θ with L_2-derivative Λ_θ, and assume that*

$$\sup_{\theta'} \mathrm{E}_{\theta'}\,|\psi_\theta|^2 < \infty \tag{1.16}$$

for all θ' s.t. $|\theta' - \theta| = \mathrm{O}(n^{-1/2})$. Then (1.15) entails

$$\mathrm{E}_\theta\,\psi_\theta\Lambda_\theta^\tau = \mathbb{I}_k \tag{1.17}$$

On the other hand, (1.15) for $\theta' = \theta$ and (1.17) imply (1.15) for all θ' s.t. $|\theta' - \theta| = \mathrm{O}(n^{-1/2})$.

It is thus no restriction to require for any influence function ψ_θ arising in (1.15) that

$$\mathrm{E}_\theta\,\psi_\theta = 0, \qquad \mathrm{E}_\theta\,\psi_\theta\Lambda_\theta^\tau = \mathbb{I}_k \tag{1.18}$$

In the shrinking neighborhood approach, except for well-definedness and its breakdown point, all asymptotic properties of an ALE (if well initialized) may be read off from its IF:

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

$$\mathrm{asVar}(S_n) = \int \psi_\theta \psi_\theta^\tau \, dF_\theta \tag{1.19}$$

(compare (Rieder, 1994, Rem. 4.2.17(b))). Remarkably, for suitably constructed ALEs, this ASV stays constant on the shrinking neighborhood (compare (Rieder, 1994, chap. 6)).

Asymptotic Bias The *gross error sensitivity* GES (compare (Hampel et al., 1986, Chapter 2.1c)) is a measure for the maximal asymptotic bias of the estimator under infinitesimal contamination:

$$\mathrm{GES} := \sup_x |\psi_\theta(x)| \tag{1.20}$$

It may be shown (cf. (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 (1.12), (1.13) is just

$$\mathrm{asBias}(S_n) = r\,\mathrm{GES} = r\,\sup_x |\psi_\theta(x)| \tag{1.21}$$

Asymptotic MSE As a consequence of the preceding two paragraphs, the (maximal, standardized) asymptotic mean squared error (MSE) attainable in the gross error model (1.12), (1.13) with starting radius r can be calculated as

$$\mathrm{asMSE}(S_n) = r^2\,\mathrm{GES}^2 + \mathrm{tr}\,\mathrm{asVar}(S_n) \tag{1.22}$$

Suitable constructions (compare (Rieder, 1994, chap. 6)) and/or uniform integrability considerations for the starting estimator (compare Ruckdeschel (2010a)) allow to interchange quantors such that asMSE also is the standardized asymptotic maximal MSE.

Remark 1.4. Using these minimax criteria asMSE, asBias defined on whole neighborhoods for defining optimally robust estimators (OMSE, MBRE), we deviate from wide-spread use in Robust Statistics to use estimators with high breakdown point (see below) which are then in a reweighting step tuned to achieve a high efficiency (say 95%) in the ideal model: We do so, simply because you cannot quantify the protection against bias you may achieve for this "insurance premium" (i.e.; the 5% efficiency loss) as this protection will vary from model to model, and in our non-invariant case even from parameter value to parameter value. This is not to say that we do not care about efficiency in the ideal model, and OMSE will prove best among the considered robust estimators in this criterion as well, but our estimators are not tuned for this, it is achieved only as a welcome side-effect. For the record: The OBRE tuned for 95% efficiency in the ideal model at $\xi = 0.7$ may drop down to 14% efficiency for sufficiently large radius, while OMSE never drops below 68% no matter what radius.

Finite Sample Breakdown Point The *asymptotic (functional) breakdown point* (*ABP*) introduced in Hampel (1968) gives the smallest radius ε at which the maximal bias of the functional on a neighborhood of this radius produces a singularity. In this paper, though, we will focus on its finite sample counterpart, the finite sample breakdown point FSBP, Donoho and Huber (1983):

Definition 1.5 ((Hampel et al., 1986, p.98)). *The* finite sample breakdown point (FSBP) ε_n^* *of the estimator* T_n *at the sample* $(x_1, ..., x_n)$ *is given by*

$$\varepsilon_n^*(T_n; x_1, ..., x_n) := \frac{1}{n} \max\{m; \max_{i_1,...,i_m} \sup_{y_1,...,y_m} |T_n(z_1, ..., z_n)| < \infty\}, \tag{1.23}$$

where the sample $(z_1, ..., z_n)$ *is obtained from the sample* $(x_1, ..., x_n)$ *by replacing the* m *data points* $x_{i_1}, ..., x_{i_m}$ *by arbitrary values* $y_1, ..., y_m$.

As argued by Davies and Gather (2005), a certain equivariance of the considered estimator under a suitable group of transformations is required to obtain meaningful upper bounds for the breakdown point. As indicated, in the GPD model, we canonically only have scale invariance, hence we should require our estimators to be scale equivariant, which is in fact true for all considered estimators in this paper, at least asymptotically. We do not use the more comprehensive (and exhaustive) group $\mathcal{G}$ of transformations $g(\theta', \theta; x)$ from (1.11), under which we also cover shape parameter ξ, as this would not lead to a meaningful notion of equivariance. Nevertheless we note that (Davies and Gather, 2005, Thm. 3.2) implies that with $n_0 = \#\{x_i = 0\}$ in the original sample,

$$\varepsilon_n^* \leq \lfloor \frac{n - n_0 + 1}{2} \rfloor / n \tag{1.24}$$

among all equivariant estimators, where equivariance may both be scale equivariance, or equivariance under the group induced by $\mathcal{G}$.

For deciding upon which procedure to take *before* having made observations, in particular for ranking procedures in a simulation study, the FSBP from Definition 1.5 has some drawbacks: It is deliberately probability-free and based on an actual sample $(x_1, ..., x_n)$, which we assume from the ideal situation for the moment. Hence its value depends on the configuration of this sample. This is desirable when checking safety of a procedure at an actual data set, but also entails that for some of the considered estimators in this paper, a generally valid value for FSBP does not exist, and the only possible lower bound will be $1/n$. To get rid of the dependence on possibly highly improbable sample configurations, but still preserving the aspect of a finite sample, we propose an expected FSBP:

Definition 1.6 (EFSBP). *For an estimator* T_n *with FSBP* $\varepsilon_n^* = \varepsilon_n^*(T; X_1, ..., X_n)$*, which is assumed measurable, we define the expected FSBP or EFSBP as*

$$\bar{\varepsilon}_n^*(T_n) := \mathrm{E}\, \varepsilon_n^*(T_n; X_1, ..., X_n) \tag{1.25}$$

where expectation is evaluated in the ideal model.

At some places, if existent, we also consider the limit

$$\bar{\varepsilon}^*(T) := \lim_{n \to \infty} \bar{\varepsilon}_n^*(T_n) \tag{1.26}$$

and which, for brevity, we also call EFSBP where clear from the context.

Weighted by their (ideal) occurrence probability, by this definition, improbable sample configurations of the ideal sample—*before* adding arbtirary contamination by replacement—are smoothed out; we still cannot exclude these configurations, but, usually, by corresponding Chebyshev-type inequalities, for growing sample size n, these will occur with decreasing probability, and ε_n^* will concentrate about $\bar\varepsilon^*$. Hence, in practice, without extra knowledge, the user can rely on being protected against up to $\bar\varepsilon_n^*(T)n$ outliers.
By averaging, EFSBP is closer again to the functional breakdown point of Hampel (1968), while still keeping the finite sample aspect of FSBP. By dominated convergence though, the limit of EFSBP will coincide with the ABP whenever the FSBP converges to the ABP.

Remark 1.7. Small values of ε_n^* for particular samples are not particular for GPD: In the one-dimensional normal scale model, we can already have FSBP of 0 for the median of absolute deviations *Mad*, if all original x_i are 0. This event (and similarly extraneous sample configurations) however occurs with probability 0, while in our case these samples can occur with small but positive ideal probability.

In this paper, EFSBP turns out useful in the context of the Pickands and MedMad-type estimators (see subsections 2.5 and 2.7 for details): In both situations, breakdown can occur if we move all observations lying in the interval $\hat I_n = (a_1\hat q_{1,n}, a_2\hat q_{2,n})$ for $0 \neq a_i \in \mathbb{R}$ and $\hat q_{i,n}$ suitable empirical quantiles outside $\hat I_n$. Now the number $\hat N_n$ of observations from the ideal sample lying in $\hat I_n$ is random, hence the $\mathrm{FSBP} = \hat N_n/n$ varies according this number, and we even have a positive, although very small probability $p_0 := P^X(\hat N_n = 0) > 0$ for breakdown already in the ideal model, i.e.; $\varepsilon_n^* = 0$, where P^X the ideal distribution.

To get hand on actual values of EFSBP and p_0, we have the following

Proposition 1.8. *Consider $\hat N_n^0$, $\hat N_n'$, $\hat N_n''$ as defined in (2.47), (2.81), (2.82) and write $\bar F$ for $1-F$. Then*

(a) *setting $i_1 = \lfloor n/2 \rfloor$, $i_2 = \lceil 3n/4 \rceil$, and abbreviating $2F^{-1}(u)$ by q_2, we obtain for $l \in \{0,\ldots,i_2 - i_1 - 1\}$*

$$P(\hat N_n^0 = l) = n\int_0^1 \binom{n-1}{i_1-1,\, i_2-i_1-l-1} u^{i_1-1}\big(F(q_2)-u\big)^{i_2-i_1-l-1}\bar F(q_2)^{n-i_2+l+1}\,du \tag{1.27}$$

(b) *using the upper median and abbreviating $(k+1)F^{-1}(u)$ by q_k, we obtain for $l \in \{0,\ldots,n/2-1\}$*

$$P(\hat N_n' = l) = n\int_0^1 \binom{n-1}{n/2,\, l} u^{n/2}\big(F(q_k)-u\big)^{l}\bar F(q_k)^{n/2-1-l}\,du \tag{1.28}$$

(c) *writing q_+ for $(1+k\check q_k)F^{-1}(u)$ and q_- for $(1-\check q_k)F^{-1}(u)$, we obtain for $l \in \{0,\ldots,n/2-1\}$*

$$\begin{aligned} P(\hat N_n'' = n/2-l) \;&=\; n\sum_{l_2=0}^{l}\binom{n-1}{n/2-l_2-1,\, l_2,\, l-l_2}\int_0^1 F(q_-)^{n/2-l_2-1}\big(u-F(q_-)\big)^{l_2}\times \\ &\qquad \times\big(F(q_+)-u\big)^{l-l_2}\big(1-F(q_+)\big)^{n/2+l_2-l}\,du \end{aligned} \tag{1.29}$$

By means of this proposition, in Table 1, we determine $\bar{\varepsilon}_n^*$ for PE, PE*, MedMad, and kMedMad ($k = 10$); apparently it is quickly converging in n, so $\bar{\varepsilon}^*$ gives indeed a useful bound on average.

estimator	$n = 40$	$n = 100$	$n = 1000$	$n = \infty$
PE	9.48%	7.61%	6.53%	6.42%
PE*	8.98%	6.85%	7.07%	7.02%
MedMad	20.41%	19.32%	18.66%	18.58%
kMedMad	29.16%	30.28%	30.94%	31.02%

Table 1 $\bar{\varepsilon}_n^*$ for PE, PE*, MedMad, and kMedMad ($k = 10$)

Again by Proposition 1.8, in Table 2, we determine p_0 for same settings.

estimator	$n = 40$	$n = 100$	$n = 1000$
PE	7.0e−02	1.3e−03	1.6e−029
PE*	5.4e−02	6.9e−04	2.6e−032
MedMad	2.7e−04	1.2e−09	5.1e−090
kMedMad	3.3e−06	6.3e−14	2.7e−126

Table 2 p_0 for PE, PE*, MedMad, and kMedMad ($k = 10$)

But, by corresponding CLT arguments, the empirical quantiles coincide with the population ones q_i up to $\mathrm{O}(n^{-1/2+\delta/2})$—except for an event with probability $\mathrm{O}(\exp(-2n^\delta))$ (Hoeffding); hence setting $I = (a_1 q_1, a_2 q_2)$, EFSBP in this context will just be $P^X(I) + \mathrm{O}(n^{-1/2+\delta/2})$, and in the limit $P^X(I)$.

To illustrate the quantity of the $\mathrm{O}(n^{-1/2+\delta/2})$-term, using the actual distribution of $\hat{N}_n$ given in Proposition 1.8 in Table 3 we determine the p_1-quantile of ε_n^* for $p_1 = 0.95^{10^{-4}}$, i.e.; the minimal number q_1, such that with probability 0.95 we will not see realizations with $\varepsilon_n^* < q_1$ in 10000 runs of sample size n; note that the minimal number of ε_n^* is $1/n$ which explains the decrease in n for PE and PE* between $n = 40$ and $n = 100$.

estimator	$n = 40$	$n = 100$	$n = 1000$	$n = \infty$
PE	2.50%	1.00%	1.30%	6.42%
PE*	2.50%	1.00%	2.60%	7.02%
MedMad	2.50%	5.00%	13.60%	18.58%
kMedMad	5.00%	15.00%	26.20%	31.02%

Table 3 q_1 for PE, PE*, MedMad, and kMedMad ($k = 10$)

1.3 Computational and Numerical Aspects

So far, we have just set the statistical framework; for an estimator to be useful in practice though, computational and numerical aspects deserve attention. In this respect, our estimator can be classified into four classes:
The first group comprises estimators which have closed-form representations and hence can be computed non-iteratively (after possibly sorting the observations, which is well known to be feasible in $\mathrm{O}(n \log(n))$ in time). In this paper this group covers PE, PE*, PicM, MME. As to computation

time, their evaluation is by magnitudes faster than the other groups, which makes them attractive for batch uses.

MLE, SMLE, and MDE are M-estimators, i.e.; obtained by optimizing a corresponding criterion function, which are solved iteratively by using R function `optim` and hence need a suitable initialization to find the "right" local optimum.

MMed, MedMad and kMedMad are zeros of corresponding (systems of) equations, hence Z-estimators. In fact we may reduce the systems to two (computationally independent) one-dimensional equations (one for determining the population Λ-Median respectively the population (k)Mad, one for solving for parameter ξ), hence in each case may use R function `uniroot` where the search interval for the (k)MedMad in case of the GPD is canonically $[0, m]$, m the population median.

Finally, MBRE and OMSE are one-step constructions, hence depend 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 a mean. The computation of the influence function at the starting estimate is not trivial, however, and to speed this up, on page 35, we present Algorithm 2.7.

As to computations, we make use of R, R Development Core Team (2009), and addon-packages `ROptEst`, Kohl and Ruckdeschel (2009), `POT`, Ribatet (2009), available on `cran.r-project.org`.

2 Estimators

In this section, for the listed estimators, we consider influence function and breakdown point and compare them as to asVar, asBias and asMSE.

2.1 Maximum Likelihood Estimator

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

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

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 with $k = 10$ as defined in subsection 2.7.2.

IF We have already seen that our model is continuously L_2-differentiable with L_2 derivative Λ_θ from (1.3) and positive definite Fisher information $\mathcal{I}_\theta$ from (1.6). The likelihood in addition is even pointwise smooth, so that for any allowed θ and for any θ_1, θ_2 in a neighborhood of θ

$$|l_{\theta_1}(x) - l_{\theta_2}(x)| \le \Lambda_\theta(x)|\theta_1 - \theta_2| \tag{2.2}$$

Thus by (van der Vaart, 1998, Thm. 5.39) (a suitably initialized version of) the MLE is an ALE with influence function

$$\mathrm{IF}_\theta(z; \mathrm{MLE}, F) = \mathcal{I}_\theta^{-1}\Lambda_\theta(z) \tag{2.3}$$

In the sequel, we call any estimator attaining this IF MLE; in particular, such a variant $\theta'_{\mathrm{MLE};n}$ may be achieved with a one-step construction

$$\theta'_{\mathrm{MLE};n} = \theta_n^{(0)} + \frac{1}{n}\sum_{i=1}^{n} \mathrm{IF}_{\theta_n^{(0)}}(z(X_i); \mathrm{MLE}, F) \tag{2.4}$$

for a starting estimator $\theta_n^{(0)}$ which is at least $n^{1/4+0}$ consistent, but may be chosen such that it is uniformly integrable in n; for details, c.f. Ruckdeschel (2010a). 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 (1.5), we obtain

$$\tilde{\psi}(v) = \mathcal{I}_\theta^{-1}\tilde{\Lambda}_\theta(v) = \frac{\xi+1}{\xi^2}\left(\begin{array}{c} -(\xi^2+\xi)\log(v) + (2\xi^2+3\xi+1)v^\xi - (\xi^2+3\xi+1) \\ \xi\log(v) - (2\xi^2+3\xi+1)v^\xi + (3\xi+1) \end{array}\right) \tag{2.5}$$

As to invariance/equivariance, we note that

$$\mathrm{IF}_{(\xi,\beta)}(x; \mathrm{MLE}, F) = d_\beta\,\mathrm{IF}_{(\xi,1)}(x/\beta; \mathrm{MLE}, F) \tag{2.6}$$

This invariance translates into at least asymptotic equivariance of the one-step construction (2.4).

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

$${\mathcal{I}_\theta}^{-1} = (1+\xi)\left(\begin{array}{cc} \xi+1 & -\beta \\ -\beta & 2\beta^2 \end{array}\right) \tag{2.7}$$

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: By replacing just one observation by some sufficiently large value, the log-term present in the shape component of the sum of the scores function Λ_ξ evaluated at the contaminated sample gets so large that only a huge value of ξ can pull back the equation to zero. One has to admit, though, that one only can approximate this breakdown for finite samples and finite contamination with really large contaminations $\sim 10^{10}$.

2.2 Skipped Maximum Likelihood Estimators

Skipped Maximum Likelihood Estimators (SMLE) as proposed in Cope et al. (2009) are ordinary MLE, skipping the largest k observations. This has to be distinguished from the better investigated *trimmed/weighted MLE* 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 the (deterministic) bias B_n, and (3) not correcting for the bias at all, but, conformal to our shrinking neighborhood setting, to let the skipping proportion α shrink at the same rate. Strategy (3) essentially models the common practice where α is often chosen small, and the bias correction is omitted. We only pursue strategy (3) in the sequel, and set $\alpha = \alpha_n = r'/\sqrt{n}$ for some r' larger than the actual r. This way indeed bias becomes asymptotically negligible:

Lemma 2.1. *Consider SMLE with skipping rate α_n. Then, in our ideal GPD model, the bias B_n of SMLE is bounded in n from above by $\bar{c} \limsup \alpha_n \log(n)$ for some constant $\bar{c} < \infty$.*
Whenever, for some $\beta \in (0, 1]$, $\liminf_n \alpha_n n^\beta > 0$, 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 $\underline{c}' > 0$.

In view of Ruckdeschel (2010a), for $\alpha_n = r'/\sqrt{n}$, this makes for an admissible starting estimator. Yet, 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, by skipping, SMLE in fact does not estimate θ but $d(\theta) = \theta + B_\theta$, B_θ the bias 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:

$$T(F) = \frac{1}{1-\alpha} \int_0^{1-\alpha} \Lambda_\theta(F^{-1}(s))ds \tag{2.8}$$

The influence function, referring to (Huber, 1981, Chapter 3.3), is analogous to the influence function of the trimmed mean:

$$\mathrm{IF}_\theta(z;\mathrm{SMLE},F) = \mathcal{I}_\theta{}^{-1} \begin{cases} \frac{1}{1-\alpha}[\Lambda_\theta(z) - W(F)], & 0 \le x \le F^{-1}(1-\alpha) \\ \frac{1}{1-\alpha}[\Lambda_\theta(F^{-1}(1-\alpha)) - W(F)], & x > F^{-1}(1-\alpha) \end{cases} \tag{2.9}$$

$$W(F) = (1-\alpha)\,\mathrm{SMLE}(F) + \alpha\Lambda_\theta(F^{-1}(1-\alpha)) \tag{2.10}$$

It enjoys the same equivariance as the MLE, i.e.

$$\mathrm{IF}_{(\xi,\beta)}(x;\mathrm{SMLE},F) = d_\beta\,\mathrm{IF}_{(\xi,1)}(x/\beta;\mathrm{SMLE},F) \tag{2.11}$$

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 As a consequence of Lemma 2.1, for a shrinking rate $\alpha_n = r'/\sqrt{n}$, asymptotic bias of SMLE is finite, but standardized by $\sqrt{n}$ is of order $\log(n)$, hence asymptotically infinite. As follows from boundedness of the IF (locally uniform in θ), the extra bias induced by contamination is of unstandardized order $\mathrm{O}(n^{-1/2})$, hence eventually dominated by B_n. Again we skip analytic terms and only include numerical values in the tables in the end.

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

2.3 Classical (first and second) moment-based estimator

Due to the fat tails of the GPD for sufficiently large scale parameter ξ, the rth moments of the GPD only exist for $\xi < 1/r$. Hence moment-based estimators only have a restricted application range. This is especially true in case of operational risk, where infinite mean models usually occur ($\xi > 1$) Neslehova et al. (2006).

In case of the GPD a classical moment-based estimator MME may be computed from empirical first and second moment. The first two theoretical moments of GPD are respectively:

$$m_1 = \frac{\beta}{1-\xi}, \qquad m_2 = \frac{2\beta^2}{(1-\xi)(1-2\xi)} \tag{2.12}$$

Hence moment-based estimators for $\xi < 0.5$ (finite second moment) can explicitely be defined as

$$\hat\xi = \frac{1}{2}\left(\frac{m_2 - 2{m_1}^2}{m_2 - {m_1}^2}\right), \quad \hat\beta = \frac{1}{2}\left(\frac{m_1 m_2}{m_2 - {m_1}^2}\right), \tag{2.13}$$

Let D be a Jacobian matrix with elements $d_{1,1}$, $d_{1,2}$, $d_{2,1}$, and $d_{2,2}$, which we obtain as

$$d_{11} = \frac{d\hat{\xi}}{dm_1} = \frac{2(\xi-1)^2(2\xi-1)}{\beta}, \qquad d_{12} = \frac{d\hat{\xi}}{dm_2} = \frac{(2\xi-1)^2(\xi-1)^2}{2\beta^2}, \tag{2.14}$$

$$d_{21} = \frac{d\hat{\beta}}{dm_1} = (4\xi-3)(\xi-1), \qquad d_{22} = \frac{d\hat{\beta}}{dm_2} = \frac{(2\xi-1)^2(\xi-1)}{2\beta}. \tag{2.15}$$

IF The influence functions of the moments are simply

$$\mathrm{IF}(x; m_1, F) = x - m_1, \quad \mathrm{IF}(x; m_2, F) = x^2 - m_2 \tag{2.16}$$

By the delta method, hence the influence function of this moment-based estimator is

$$\mathrm{IF}(x; \mathrm{MME}, F) = D\big(\mathrm{IF}(x; m_1, F), \mathrm{IF}(x; m_2, F)\big)^{\tau} \tag{2.17}$$

It enjoys the same equivariance as the MLE, i.e.

$$\mathrm{IF}_{(\xi,\beta)}(x; \mathrm{MME}, F) = d_\beta \, \mathrm{IF}_{(\xi,1)}(x/\beta; \mathrm{MME}, F) \tag{2.18}$$

ASB Both coordinates of the influence function of MME are parabolas in x, hence unbounded, so the asymptotic bias is infinite.

ASV Asymptotic normality requires $\xi < 0.25$ (finite fourth moment). The asymptotic variance of the moment-based estimators then is $V = D\Sigma D^T$ where Σ is a covariance matrix of m_1 and m_2 with elements

$$\sigma_{11} = \frac{\beta^2}{(1-\xi)^2(1-2\xi)}, \quad \sigma_{12} = \frac{4\beta^3}{(1-\xi)^2(1-2\xi)(1-3\xi)},$$
$$\sigma_{22} = \frac{4\beta^4(5-11\xi)}{(1-\xi)^2(1-2\xi)^2(1-3\xi)(1-4\xi)} \tag{2.19}$$

Combining these, we obtain:

$$\mathrm{asVar}(\mathrm{MME}) = \frac{(1-\xi)^2}{(1-4\,\xi)\,(1-3\,\xi)} \begin{pmatrix} V_{1,1}, & V_{1,2} \\ V_{1,2}, & V_{2,2} \end{pmatrix} \tag{2.20}$$

for

$$V_{1,1} = (1-2\,\xi)(1-\xi+1+6\,\xi^2), \quad V_{1,2} = -\beta\,(1-4\,\xi+12\,\xi^2),$$
$$V_{2,2} = 2\,\frac{\beta^2(1-7\,\xi+18\,\xi^2-12\,\xi^3)}{1-2\,\xi} \tag{2.21}$$

FSBP Letting one replaced observation x_0 tend to ∞, for the empirical moments $\hat{m}_1$ and $\hat{m}_2$ we get $\hat{m}_1/x_0 \to 1$, $\hat{m}_2/x_0^2 \to 1$, hence, as the corresponding denominator by Cauchy-Schwartz never becomes negative, $\hat{\xi} \to -\infty$, which shows that the FSBP of MME is $1/n$.

2.4 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 hence, as for MLE and SMLE, depends on a suitable initialization. We use Cramér-von-Mises distance defined for cumulative distribution functions (c.d.f.'s) F, G and some σ-finite measure ν on $\mathbb{B}^k$ as

$$d_{\mathrm{CvM}}(F,G)^2 = \int (F(x) - G(x))^2\, \nu(dx) \tag{2.22}$$

i.e., by MDE we denote

$$\mathrm{MDE} = \operatorname*{argmin}_{\theta} d_{\mathrm{CvM}}(\hat{F}_n, F_\theta) \tag{2.23}$$

In this paper, we use $\nu = P_\theta$; another setting common in the literature uses the empirical, $\nu = \hat{P}_n$. As initialization we again use Hybr from subsection 2.7.2. MDE is known to have good global robustness properties: it is asymptotically linear ((Rieder, 1994, Remark 6.3.9(a))) with bounded IF—bounded by $\mathrm{E}_\theta\, |\mathcal{J}_\theta^{-1}\Delta_\theta|^2$ ((Rieder, 1994, 4.2 eq. (55)))—and, according to Donoho and Liu (1988a), upto a factor 2 achieves the smallest sensitivity to contamination among Fisher-consistent[2] estimators.

Remark 2.2. Another possible distance with the same property would be Kolmogorov distance $d_\kappa(F,G) = \sup_x |F(x) - G(x)|$. As shown in Donoho and Liu (1988b), however, the corresponding minimum distance estimator has a non-stable variance on arbitrarily small neighborhoods; in addition, it is not asymptotically linear; hence, in this paper, we have not considered it more closely.

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

$$\begin{aligned} \mathrm{IF}(x;\mathrm{MDE},F) &= \mathcal{J}_\theta^{-1}\Big(-\int_0^x \Delta_\theta(y)\,F(dy) + \int_0^\infty (1-F(y))\Delta_\theta(y)\,F(dy)\Big) \\ &=: \mathcal{J}_\theta{}^{-1}(\tilde{\varphi}_\xi(x), \tilde{\varphi}_\beta(x)) \end{aligned} \tag{2.24}$$

where Δ_θ is CvM derivative and $\mathcal{J}_\theta$ is the CvM Fisher information as defined, e.g. in (Rieder, 1994, Definition 2.3.11)): The CvM derivative for GPD is obtained as derivative of the c.d.f. w.r.t. the parameters: $\Delta_\theta = (\Delta_\xi, \Delta_\beta)^T$ with

$$\Delta_\xi(z) = -\frac{1}{\xi^2}\,(1+\xi z)^{-\frac{1}{\xi}} \log(1+\xi z) + \frac{z}{\xi}\,(1+\xi z)^{-\frac{1}{\xi}-1} \tag{2.25}$$

$$\Delta_\beta(z) = -\frac{z}{\beta}\,(1+\xi z)^{-\frac{1}{\xi}-1} \tag{2.26}$$

[2] Recall that an estimator T is Fisher-consistent if $T(F_\theta) = \theta$ for all parameter values θ; Fisher-consistency of CvM-MDE in turn is implied by local identifiability (i.e.; regular Fisher information) and L_2-differentiability of the model (compare (Rieder, 1994, Lem. 6.3.3))

and the CvM Fisher information is obtained as $\mathcal{J}_\theta = \int \Delta_\theta \Delta_\theta^T dF$, the inverse of which in case of the GPD model is:

$$\mathcal{J}_\theta{}^{-1} = 3(\xi+3)^2 \begin{pmatrix} \frac{18(\xi+3)}{(2\xi+9)} & -3\beta \\ -3\beta & 2\beta^2 \end{pmatrix} \tag{2.27}$$

Hence, using again $v^{-\xi}(z) = 1 + \xi z$ as in (1.5),

$$\tilde{\varphi}_\xi(v(z)) = \frac{19+5\xi}{36(3+\xi)(2+\xi)} + \frac{1}{\xi}v^2\log(v) + \frac{2-\xi}{4\xi^2}v^2 - \frac{1}{\xi^2(2+\xi)}v^{2+\xi} \tag{2.28}$$

$$\tilde{\varphi}_\beta(v(z)) = \frac{5+\xi}{6(3+\xi)(2+\xi)\beta} - \frac{1}{2\xi\beta}v^2 + \frac{1}{\xi\beta(2+\xi)}v^{2+\xi} \tag{2.29}$$

Apparently the same invariance/equivariance as for MLE, SMLE, and MME is present here as well.

Remark 2.3. The fact that MDE is asymptotically linear with the IF just given allows for an alternative to the numerical minimization of the distance: As indicated in case of the MLE in (2.1), 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. Analytic terms are rational functions in ξ and β; for the interested reader we have `MAPLE` scripts to determine it. The actual analytic terms are as follows[3]:

$$\mathrm{asVar}(\mathrm{MDE}) = \frac{(3+\xi)^2}{125\,(5+2\,\xi)\,(5+\xi)^2} \begin{pmatrix} V_{1,1}, & V_{1,2} \\ V_{1,2}, & V_{2,2} \end{pmatrix} \tag{2.30}$$

for

$$V_{1,1} = 81\left(16\,\xi^5 + 272\,\xi^4 + 1694\,\xi^3 + 4853\,\xi^2 + 7276\,\xi + 6245\right)(2\,\xi+9)^{-2}, \tag{2.31}$$

$$V_{1,2} = -9\beta\left(4\,\xi^4 + 86\,\xi^3 + 648\,\xi^2 + 2623\,\xi + 4535\right)(2\,\xi+9)^{-1}, \tag{2.32}$$

$$V_{2,2} = \beta^2\left(26\,\xi^3 + 601\,\xi^2 + 3154\,\xi + 5255\right) \tag{2.33}$$

ASB The IF of the CvM MDE is known to be bounded Rieder (1994), so ASB is finite. For our reference parameter value, we have determined it numerically in Table 6.

FSBP Due to the lack of invariance in the GP situation, (Donoho and Liu, 1988a, 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.

[3] The expressions given in Linde (2007) obviously contain an error in the first paragraph of page 150: only involving linear and square terms of expressions with exponent $-2/\xi$, expressions with an exponent of $-3/\xi$ cannot arise; this error then is propagated until the final expressions of asVar in the cited reference.

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 $V_i = (1 + \frac{\xi}{\beta} X_i)^{-1/\xi}$,

$$\sum_i \xi \tilde{\varphi}_\xi(V_i; \hat{\xi}) = 0, \qquad \sum_i \beta \tilde{\varphi}_\beta(V_i; \hat{\xi}) = 0 \tag{2.34}$$

where we multiply the equations by ξ and β respectively, to avoid singularities in $\xi = 0$, $\beta = 0$. Now by placing m observations on the respective suprema/infima of the coordinates of $\tilde{\varphi}$, we see that we can no longer pull back the sum to 0 (not even if all remaining ideal observations were placed at the respective infimum/supremum), once $m \sup_v \sup_\xi \varphi_{\cdot} > -(n-m) \inf_v \inf_\xi \varphi_{\cdot}$, respectively $m \inf_v \inf_\xi \varphi_{\cdot} < -(n-m) \sup_v \sup_\xi \varphi_{\cdot}$, so that

$$\varepsilon_n^* \leq \min \Big\{ \frac{-\inf_v \inf_\xi \varphi_{\cdot}}{\sup_v \sup_\xi \varphi_{\cdot} - \inf_v \inf_\xi \varphi_{\cdot}}, \frac{\sup_v \sup_\xi \varphi_{\cdot}}{\sup_v \sup_\xi \varphi_{\cdot} - \inf_v \inf_\xi \varphi_{\cdot}}, \quad \cdot = \xi, \beta \Big\} \tag{2.35}$$

Except for the optimization in ξ, this nothing but the formula given in (Huber, 1981, Chap. 3, eqs. (2.39) and (2.40)), although, to make the inequality in (2.35) an equality, we would need to show that we cannot produce a breakdown with less than this bound, which we do not see how to. Evaluating bound (2.35) numerically, this gives a value of $4/9 \doteq 36.37\%$, which is achieved for $v = 0$ (and $\xi \to 0$) or, equivalently, letting the m replacing observations tend to ∞.

To see how realistic this value is, we determine the FSBP empirically by simulations: On each of $M = 100$ samples of size $n = 1000$ observations from a GPD with $\xi = 0.7$, $\beta = 1$, we have replaced m observations, for $m = 1, \ldots, 400$ by 10^{10} and subsequently evaluated MDE. In Figure 1, we produce an empirical max-bias-curve, plotting $m/1000$ against the corresponding empirical bias. We see that there is an extremal steep increase at about 0.354, so we conjecture that (E)FSBP should be approximately equal to this value; however, we should note that MDE needs an initialization, which, too, must not be broken down, and that, so far, we have not found any possible initialization with (E)FSBP larger than 0.346.

2.5 Pickands Estimator and PE-type Estimators

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). Pickands estimators for ξ and β is defined as

$$\hat{\xi} = \frac{1}{\log(2)} \log \frac{M_4 - M_2}{M_2}, \quad \hat{\beta} = \hat{\xi} \frac{M_2^2}{M_4 - 2M_2} \tag{2.36}$$

Looking more closely at the construction of PE, we note that this technique is not limited to 50% and 75% quantiles. More specifically, let $a > 1$ and consider the emprical α_i-quantiles for $\alpha_1 = 1 - 1/a$ and $\alpha_2 = 1 - 1/a^2$ denoted for the ease of comparison with the original PE 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

$$\hat{\xi} = \frac{1}{\log a} \log \frac{M_4(a) - M_2(a)}{M_2(a)}, \quad \hat{\beta} = \hat{\xi} \frac{M_2(a)^2}{M_4(a) - 2M_2(a)} \tag{2.37}$$

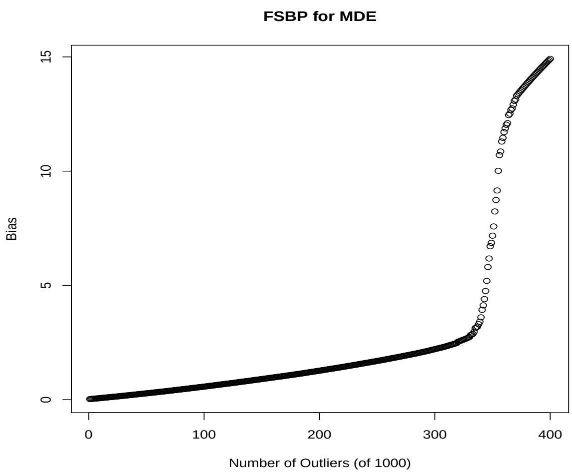


Figure 1 Empirical Bias for FSBP of MDE CvM

Apparently for any $a > 1$, PE(a) enjoys the corresponding equivariance as MLE, SMLE, MME, 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 $h_i,\ i = 1, ..., k$ may be read off from (Rieder, 1994, Chapter 1.5) and gives

$$\mathrm{IF}(x; T_L, F) = \sum_{i=1}^{k} h_i \frac{\alpha_i - \mathbb{I}(x \le F^{-1}(\alpha_i))}{f(F^{-1}(\alpha_i))} \tag{2.38}$$

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

$$\mathrm{IF}_\xi(x; \mathrm{PE}(a), F) = \sum_{i=1}^{2} h_{\xi,i}(a) \frac{\alpha_i(a) - \mathbb{I}(x \le M_{2i}(a))}{f(M_{2i}(a))}, \tag{2.39}$$

$$\mathrm{IF}_\beta(x; \mathrm{PE}(a), F) = \sum_{i=1}^{2} h_{\beta,i}(a) \frac{\alpha_i(a) - \mathbb{I}(x \le M_{2i}(a))}{f(M_{2i}(a))} \tag{2.40}$$

with weights

$$h_{\xi,1}(a) = -\frac{1}{\log(a)} \frac{M_4}{M_2(M_4 - M_2)}, \quad h_{\xi,2}(a) = \frac{1}{\log(a)} \frac{1}{M_4 - M_2} \tag{2.41}$$

$$h_{\beta,1}(a) = h_{\xi,1}(a) \frac{(M_2)^2}{M_4 - 2M_2} + \frac{1}{\log(a)} \frac{2M_2(M_4 - M_2)}{(M_4 - 2M_2)^2} \log \frac{M_4 - M_2}{M_2} \tag{2.42}$$

$$h_{\beta,2}(a) = h_{\xi,2}(a) \frac{(M_2)^2}{M_4 - 2M_2} - \frac{1}{\log(a)} \frac{(M_2)^2}{(M_4 - 2M_2)^2} \log \frac{M_4 - M_2}{M_2} \tag{2.43}$$

where $M_{2i} = M_{2i}(a)$, $i = 1, 2$. Apparently we have again equivariance,

$$\mathrm{IF}_{(\xi,\beta)}(x; \mathrm{PE}(a), F) = d_\beta \, \mathrm{IF}_{(\xi,1)}(x/\beta; \mathrm{PE}(a), F) \tag{2.44}$$

ASV Abbreviating $\alpha_i(a)$ by α_i and $1-\alpha_i$ by $\bar\alpha_i$, the asymptotic covariance for PE(a) is

$$\mathrm{asVar}(\mathrm{PE}(a)) = D(a)^T \Sigma(a) D(a), \tag{2.45}$$

$$\Sigma(a) = \beta^2 \left(\begin{array}{cc} \alpha_1 \bar\alpha_1^{-1-2\xi} & \alpha_1 \bar\alpha_1^{-1-\xi} \bar\alpha_2^{-\xi} \\ \alpha_1 \bar\alpha_1^{-1-\xi} \bar\alpha_2^{-\xi} & \alpha_2 \bar\alpha_2^{-1-2\xi} \end{array}\right), \quad D(a) = \left(\begin{array}{cc} h_{\xi,1}(a) & h_{\xi,2}(a) \\ h_{\beta,1}(a) & h_{\beta,2}(a) \end{array}\right) \tag{2.46}$$

ASB The IF of PE(a) is bounded, so the ASB is also finite; it is computed numerically for the reference parameter value.

FSBP Apparently, we can render the scale estimator arbitrarily large for $M_4(a)$ sufficiently large, so $\varepsilon_n^* < 1 - \alpha_2(a) = 1/a^2$; also, if $\mu = 0$, $M_2(a) = 0 + 0$ has the same effect, so in this case, $\varepsilon_n^* < \alpha_1(a) = 1 - 1/a$. No matter the value of μ, the denominator of $\hat\beta$, $M_4(a) - 2M_2(a)$ may cause problems, yielding negative $\hat\beta$, once $M_4(a) \le 2M_2(a)$, which certainly happens if, in an ideally distributed sample, we replace all observations X_i, $2M_2(a) \le X_i \le M_4(a)$ by $M_2(a)$, so

$$n\varepsilon_n^* \le \hat N_n^0 := \#\{X_i \,\big|\, 2M_2(a) \le X_i \le M_4(a)\} \tag{2.47}$$

As for the respective population quantiles, we clearly have $M_4^\natural(a) > 2M_2^\natural(a)$ so by the strong law of large numbers, in the ideal situation it holds that $M_4(a) > 2M_2(a)$ eventually in n almost surely. Hence, by the Hoeffding inequality for quantiles, up to an event of exponentially small probability, $\varepsilon_n^* = \pi_\xi + \mathrm{O}_{P_\theta^n}(n^{-1/2+\delta})$, where

$$\pi_\xi = P_\theta(2M_2^\natural < X_1 \le M_4^\natural) = (2^{\xi+1} - 1)^{-1/\xi} - 1/4 \tag{2.48}$$

and we obtain

$$\bar\varepsilon^* = \bar\varepsilon^*(a) = \min\{\pi_\xi(a), 1/a^2\} \tag{2.49}$$

2.5.1 PE-type estimator tuned for high EFSBP

If we want to tune for maximal EFSBP within the class of PE(a) estimators, we have to maximize $\varepsilon_n^*(a)$ for $a > 1$, which can be done numerically, and in case of our reference parameter $\xi = 0.7$ gives $a^* = 2.658$ with a EFSBP of 7.02%; for the sequel, denote this estimator by PE*; also note that for the classical PE, we obtain $\bar\varepsilon^*(a = 2) \doteq 6.42\%$; for the figures for $\bar\varepsilon_n^*$, for $n = 40, 100, 1000$ see Table 1.

2.5.2 PE-type estimator tuned for low(er) variance

Although easy to compute and with acceptable robustness properties, for finite sample sizes, both PE, and, a little better, PE* come with high asymptotic covariances. To improve upon this, we introduce a PE-type estimator PicM by averaging variants PE(a) for different values of a, i.e., PicM is the arithmetic mean of PE(a_j) for a_j, $j = 1, \ldots, 15$ equally spaced in $(2, 2.5)$. This approach is similar to EPM, but in addition tries to find a tradeoff between FSBP and variance.

IF PicM is just an exact linear combination of Pickands-type estimators, so

$$\mathrm{IF}(x; \mathrm{PicM}, F) = \frac{1}{15} \sum_{j=1}^{15} \mathrm{IF}_{\xi}(x; \mathrm{PE}(a_j), F) \tag{2.50}$$

ASV Denote $H \in \mathbb{R}^{2\times 30}$ the matrix filled with row $h_{\xi,i}(a_j)$, and row $h_{\beta,i}(a_j)$, each for $i = 1, 2$, $j = 1, \ldots, 15$, and Σ the common covariance matrix of all quantiles $M_{2i}(a_j)$ with entry $(i_1, j_1; i_2, j_2)$ given by

$$\Sigma_{i_1,j_1;i_2,j_2} = \frac{\min\big(\alpha_{i_1}(a_{j_1}), \alpha_{i_2}(a_{j_2})\big)(1 - \max(\alpha_{i_1}(a_{j_1}), \alpha_{i_2}(a_{j_2})))}{f(M_{2i_1}(a_{j_1}))f(M_{2i_2}(a_{j_2}))} \tag{2.51}$$

Then the asymptotic covariance is $V = H\Sigma H^{\tau}$.

ASB Again, the IF of PicM is bounded, so the ASB is also finite; it is computed numerically for our reference parameter in Table 6.

FSBP From the discussion of the general EFSBP for PE(a), it is clear that the breakdown point of this estimator cannot be better than the worst of all its components, being the classical PE in our case; on the other hand, at least one of the constituents has to break down for a breakdown of PicM, and for this we have to replace $\pi_{\xi}(a_i)n$ observations, which is easiest for PE, hence $\bar{\varepsilon}^* = 6.42\%$. Notice, though that the variance of PicM is smaller than that of PE due to averaging.

2.6 Method of Median Estimator

The Method of Median 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

$$\mathrm{Median}(X_i)/\beta = F_{1,\xi}^{-1}(1/2) = (2^{\xi} - 1)/\xi =: m_{\xi} \tag{2.52}$$

$$\mathrm{Median}\left(\frac{\log(1 + \frac{\xi}{\beta}X_i)}{\beta^2} - \frac{(1+\xi)X_i}{\beta\xi + \xi^2 X_i}\right) = z(\xi) \tag{2.53}$$

where $z(\xi)$ is the population median of the second (shape) coordinate of $\Lambda_{1,\xi}(X)$ for $X \sim \mathrm{GPD}(1, \xi)$ As we can solve the first equation for β and plug 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 influence function of MMed is then a linear combination of the influence function of the median of the X_i which we already have used in the PE, and the influence function of the median 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 $\Lambda_{1,\xi;2}(X) \le z(\xi)$ is of form $[q_1(\xi), q_2(\xi)]$, so that

$$\mathrm{IF}(x; \Lambda\text{-Med}, F) = \frac{\mathbb{I}(q_1 \le x \le q_2) - 1/2}{f_\theta(q_2)/l_2 - f_\theta(q_1)/l_1} \tag{2.54}$$

where

$$l_i := \frac{\partial}{\partial x}\Lambda_{1,\xi;2}(q_i) \tag{2.55}$$

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

$$\mathrm{IF}(x; \mathrm{MMed}, F) = D(\mathrm{IF}(x; \mathrm{Median}, F), \mathrm{IF}(x; \Lambda\text{-Med}, F))^\tau \tag{2.56}$$

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

$$\tilde{D}^{-1} = \mathrm{E}_\theta\, \eta_\theta \Lambda_\theta^\tau \qquad \text{for} \quad \eta_\theta(x) = \Big(\mathbb{I}(x \le m_\xi) - 1/2, \mathbb{I}(q_1 \le x \le q_2) - 1/2\Big)^\tau \tag{2.57}$$

and then to write

$$\mathrm{IF}(x; \mathrm{MMed}, F) = \tilde{D}\eta_\theta \tag{2.58}$$

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

ASV Similarly, we obtain

$$\mathrm{asVar}(\mathrm{MMed}) = \tilde{D}\Sigma(a)\tilde{D}^\tau, \qquad \Sigma(a) = \frac{1}{4}\begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix}, \quad c = 1 - 4F(q_1) \tag{2.59}$$

ASB The IF of MMed is bounded, so the ASB is also finite; it is computed numerically for the special cases looked at in detail.

FSBP The authors did not succeed to find an analytic value for neither the asymptotic nor the finite sample breakdown point. 50% by equivariance is an upper bound, though; the high frequency of failures in the simulation study for small sample sizes however indicates that FSBP should be considerably smaller; a similar study for the empirical maxBias as the one for MDE gives that for sample size $n = 40$, from a rate of outliers of $\varepsilon = 42.5\%$ on, for $n = 100$ from $\varepsilon = 35.0\%$, for $n = 1000$ from $\varepsilon = 25.1\%$, and for $n = 10000$ from $\varepsilon = 20.1\%$ on, we have but failures in solving for MMed. So we conjecture that the asymptotic breakdown point $\varepsilon^* \le 20\%$.

2.7 MedMad Estimator

Empirical median ($\hat{m}$) and median of absolute deviations ($\hat{M}$) are well known for their high breakdown point, jointly achieving the highest possible breakdown point of 0.5 among all affine equivariant estimators at symmetric, continuous distributions on the real line.

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 the MedMad estimator.

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 Mad, which has to be solved numerically as unique root M of $f_{m,\xi,\beta}(M)$ for

$$f_{m,\xi,\beta}(M) = -\left(1 + \xi\frac{M+m}{\beta}\right)^{-\frac{1}{\xi}} + \left(1 + \xi\frac{(-M+m)_+}{\beta}\right)^{-\frac{1}{\xi}} - \frac{1}{2} \tag{2.60}$$

Note that $f_{m,\xi,\beta}(M) > 0$ for $M > m$, hence the population Mad $M(\xi, \beta)$ in the GPD must always be smaller than its median, or $M(\xi, \beta)/m(\xi, \beta) \le 1$.

Now, as generally true for scale estimators, Mad $M(\xi, \beta) = \beta M(\xi, 1)$, and the empirical Mad $\hat{M}$ is scale-equivariant, i.e., $\hat{M}(\beta x_1, \ldots \beta x_n) = \beta \hat{M}(x_1, \ldots x_n)$.

The same relations hold for the median, too; hence both the quotient $q(\xi) := M(\xi, \beta)/m(\xi, \beta)$ and and its empirical counter part $\hat{q}_n$ are scale-free; so we have reduced the problem to $\beta = 1$.

Plotting the function $\xi \mapsto q(\xi)$, we see that there is a second restriction of the same sort as that $q(\xi) < 1$, induced by the fact that for all $\xi > 0$,

$$q(\xi) \ge \lim_{\xi \to 0} q(\xi) =: \check{q} \tag{2.61}$$

This function is plotted in Figure 2.

Hence matching $\hat{q}_n$ against $q(\xi)$ amounts to finding a zero $\hat{\xi}_n$ of $G(\xi) = q(\xi) - \hat{q}_n$ in the interval $(\check{q}; 1)$ 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

$$\hat{\beta}_n = \hat{m}/m(\hat{\xi}_n, 1) \tag{2.62}$$

so by construction MedMad is equivariant in the sense of (1.8).

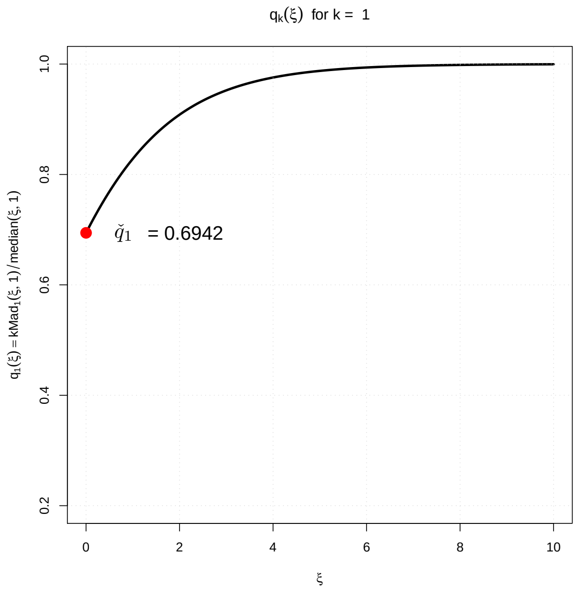


Figure 2 Function $\xi \mapsto q(\xi) =: q_{k=1}(\xi)$

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

$$G\left((\xi, \beta); (M, m)\right) = \{G^{(1)}, G^{(2)}\}^{\tau} = \left\{ \begin{array}{c} f_{m,\xi,\beta}(M) \\ \beta \frac{2^{\xi}-1}{\xi} - m \end{array} \right\} \tag{2.63}$$

By the implicit function theorem, we obtain the following matrix D to be used in the Delta method:

$$D = -\left(\frac{\partial G}{\partial(\xi, \beta)}\right)^{-1} \frac{\partial G}{\partial(M, m)} \tag{2.64}$$

Then the influence function of MedMad estimator is

$$\mathrm{IF}(x; \mathrm{MedMad}, F) = D(\mathrm{IF}(x; \mathrm{Mad}, F), \mathrm{IF}(x; \mathrm{Median}, F))^{\tau} \tag{2.65}$$

where the influence functions of median and Mad can be found in (Rieder, 1994, Chapter 1.5):

$$\mathrm{IF}(x; m, F) = \frac{\frac{1}{2} - \mathbb{I}(x \leq m)}{f(m)} \tag{2.66}$$

$$\mathrm{IF}(x; M, F) = \frac{\frac{1}{2} - \mathbb{I}(|x - m| \leq M)}{f(m + M) - f(m - M)} + \frac{f(m + M) - f(m - M)}{f(m + M) + f(m - M)} \frac{\mathbb{I}(x \leq m) - \frac{1}{2}}{f(m)} \tag{2.67}$$

while for the entries of D, abbreviating

$$v_{+} := \left(1 + \xi \frac{M + m}{\beta}\right)^{-\frac{1}{\xi}}, \qquad v_{-} := \left(1 + \xi \frac{-M + m}{\beta}\right)^{-\frac{1}{\xi}} \tag{2.68}$$

we note that

$$\frac{\partial G^{(1)}}{\partial \xi} = -v\left(\frac{v^{\xi}-1}{\xi^2} - \frac{1}{\xi}\log(v)\right)\Bigg|_{v=v_-}^{v_+}, \quad \frac{\partial G^{(1)}}{\partial \beta} = \frac{v}{\xi\beta^2}(v^{\xi}-1)\Bigg|_{v=v_-}^{v_+}, \tag{2.69}$$

$$\frac{\partial G^{(1)}}{\partial M} = \frac{1}{\beta}(v_+^{\xi+1} + v_-^{\xi+1}), \quad \frac{\partial G^{(1)}}{\partial m} = \frac{v^{\xi+1}}{\beta}\Bigg|_{v=v_-}^{v_+} \tag{2.70}$$

$$\frac{\partial G^{(2)}}{\partial \xi} = \frac{\beta}{\xi}\left(2^{\xi}\log(2) - \frac{2^{\xi}-1}{\xi}\right), \quad \frac{\partial G^{(2)}}{\partial \beta} = \frac{2^{\xi}-1}{\xi},$$

$$\frac{\partial G^{(2)}}{\partial M} = 0, \quad \frac{\partial G^{(2)}}{\partial m} = -1 \tag{2.71}$$

Again, we have equivariance,

$$\mathrm{IF}_{(\xi,\beta)}(x; \mathrm{MedMad}, F) = d_{\beta}\,\mathrm{IF}_{(\xi,1)}(x/\beta; \mathrm{MedMad}, F) \tag{2.72}$$

ASV The asymptotic covariance of the MedMad estimator is

$$\mathrm{asVar}(T) = D^T \Sigma D, \qquad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \tag{2.73}$$

Σ is covariance of the joint distribution of median and Mad with elements Serfling and Mazumder (2009):

$$a = f(m-M) + f(m+M), \qquad b = f(m-M) - f(m+M), \tag{2.74}$$

$$c = f(m-M) + f(m+M), \qquad d = b^2 + 4(1-a)bf(m), \tag{2.75}$$

$$\sigma_{11} = (4f(m))^{-2}, \quad \sigma_{12} = \sigma_{21} = (4f(m)c)^{-1}\left(1 - 4F(m-M) + \frac{b}{f(m)}\right),$$

$$\sigma_{22} = \frac{f(m)^2}{4c^2(f(m)^2 + d)} \tag{2.76}$$

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

FSBP The FSBP of 50% of the median obviously is an upper bound, implying that you could at least drive one of the parameters β and ξ to ∞. However, similarly to the Weibull case of Boudt et al. (2010), breakdown is not only entailed by moving mass to 0 or ∞, and the actual breakdown point of MedMad is smaller:

As we have seen, within the GPD model, can no longer be solved, once the quotient $\hat{q}_n$ no longer falls into $[\check{q}, 1)$; which could be achieved by either moving all observations $\hat{m} < X_i \le \hat{m} + \hat{M}$ to $2\hat{m}$

(entailing $\hat{q}_n > 1$) or by moving observations to the interval $[(1-\check{q})\hat{m}, (\check{q}+1)\hat{m}]$ up to the point that it contains $n/2$ observations (entailing $\hat{q}_n < \check{q}$) . It turns out that the first alternative amounts to moving less observations. On first glance, this would make for a "definition breakdown", but if we move the observations to $2\hat{m} - 0$, we obtain as estimator $\hat{\beta} = 0 + 0$ and $\hat{\xi} = \infty$, hence a breakdown in the original sense.

Thus, upto remainder terms of order $\mathrm{O}(n^{-1/2})$, the EFSBP of MedMad is just

$$\bar{\varepsilon}^* = \bar{\varepsilon}^*(\xi) = 1/2 - (2^{\xi+1} - 1)^{-1/\xi} \tag{2.77}$$

which for our reference parameter $\xi = 0.7$ is 18.58%; for the figures for $\bar{\varepsilon}_n^*$, for $n = 40, 100, 1000$ see Table 1. Hence contrary to Boudt et al. (2010), not only is our FSBP varying from sample to sample, but also the EFSBP depends on ξ.

2.7.1 kMedMad

The value $\bar{\varepsilon}^* = 18.58\%$ is disappointingly small, in particular if we account for the potential downward correction by the $\mathrm{O}(n^{-1/2})$ term. It is a consequence of the asymmetry present in GPD. A remedy could be to define asymmetric scale estimators about the median as follows: For a distribution F on $\mathbb{R}$ with median m let us define for $k > 0$

$$\mathrm{kMad}(F, k) := \inf\left\{\, t > 0 \,\middle|\, F(m + kt) - F(m - t) \geq 1/2 \,\right\} \tag{2.78}$$

where k in our case is chosen to be a suitable number larger than 1. Up to this modification, we may proceed as in the preceding section, i.e.; match the empirical median and kMad by the model counterparts and define the matching parameter as estimator.

IF The resulting estimator is again an ALE with IF just analogous to the one of MedMad. We only give the necessary substitutions here: In (2.67), the first indicator becomes $\mathbb{I}(-M \leq x - m \leq kM)$, and we have to substitute expressions $m+M$ by $m+kM$ in (2.68). By the chain rule, an extra factor k appears in the denominator $f(m+M)+f(m-M)$ of (2.67) (which gets $kf(m+kM)+f(m-M)$), the same in c in (2.75), and the first summand of $\partial G^{(1)}/\partial M$ in (2.70) is multiplied by k.

ASB The IF of the kMedMad estimator again is bounded, so the asymptotic bias is finite.

FSBP With respect to the MedMad estimator, for $k > 1$, we achieve higher breakdown points: Paralleling the case of the MedMad, the quotient $\hat{q}_{k;n} = \mathrm{kMad}/\hat{m}$ for a GPD must lie in the interval

$$I(k) = \big[\check{q}_k; 1\big), \qquad \check{q}_k = \lim_{\xi \to 0} q_k(\xi) \tag{2.79}$$

for $q_k(\xi) = M_k(\xi, 1)/m(\xi, 1)$.

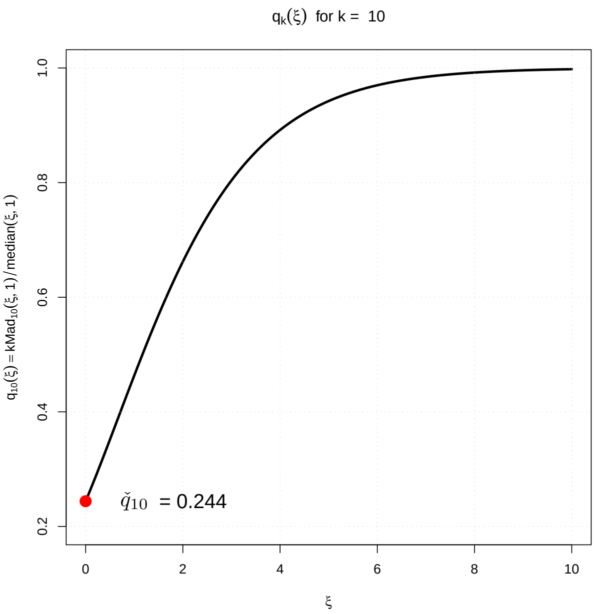


Figure 3 Function $\xi \mapsto q_{k=10}(\xi)$

This function is plotted in Figure 3 for $k = 10$.

So again two strategies can be used to produce a breakdown, i.e. either to move all observations from the right part of the interval about the median containing 50% of the observations to the maximal attainable point of $(k+1)\hat{m}$, which obviously for $k > 1$ is much harder than for $k = 1$, or to move additional observations into the interval $[(1-\check{q}(k))\hat{m}, (k\check{q}(k)+1)\hat{m}]$ until it contains $n/2$ observations. The actual FSBP is then given by the alternative needing to move less observations. More precisely

$$\bar{\varepsilon}_n^* = \min(\hat{N}_n', \hat{N}_n'')/n, \tag{2.80}$$

$$\hat{N}_n' = \#\{X_i \,|\hat{m} < X_i \le (k+1)\hat{m}\,\}, \tag{2.81}$$

$$\hat{N}_n'' = \lceil n/2 \rceil - \#\{X_i \,|\, (1-\check{q}_k)\hat{m} \le X_i \le (k\check{q}_k+1)\hat{m}\} \tag{2.82}$$

Hence, by the usual LLN arguments,

$$\bar{\varepsilon}^* = \min\big(F_\theta((k+1)m) - 1/2,\; F_\theta\big((k\check{q}_k+1)m\big) - F_\theta\big((1-\check{q}_k)m\big) - 1/2\big) \tag{2.83}$$

As to the choice of k, it turns out that a value of $k = 10$ gives reasonable values of EFSBP, asVar, asBias for a wide range of parameters ξ, as documented in Table 4. In the sequel this will be our reference value for k; for the figures for $\bar{\varepsilon}_n^*$, for $n = 40, 100, 1000$ see Table 1.

Optimizing within the class of kMedMad estimators, i.e. for varying k, with respect to the other robustness criteria for $\xi = 0.7$, we obtain Table 5, the entries of which should be compared to those of Table 6.

ξ	GES	$\mathrm{tr\,asVar}$	asMSE	ABP
0.01	4.09	12.08	16.26	0.249
0.10	3.83	10.90	14.58	0.259
0.70	4.38	12.80	17.60	0.310
1.50	5.85	19.50	28.06	0.355
4.00	10.58	52.90	80.90	0.221

Table 4: Robustness properties of kMedMad for $k = 10$ and several shape parameters compared to corresponding optimal values, i.e., MBRE (GES), MLE ($\mathrm{tr\,asVar}$), OMSE (asMSE), $\mathrm{kMedMad}(k^{\mathrm{ABP}})$, $k^{\mathrm{ABP}} = \mathrm{argmax}_k\,\mathrm{ABP}(\mathrm{kMedMad}(k))$ (ABP)

criterion	k	GES	$\mathrm{tr\,asVar}$	asMSE	ABP
(MedMad)	1.00	8.64	29.55	48.22	0.186
($k = 10$)	10.00	4.38	12.80	17.60	0.310
GES	5.47	3.81	14.50	18.13	0.328
asVar	52.77	7.21	11.31	24.32	0.259
asMSE	9.18	4.29	12.99	17.59	0.313
ABP	3.23	4.61	16.98	22.30	0.342

Table 5: Robustness properties of kMedMad for $\xi = 0.7$ and several distinguished values of k

Remark 2.4. We should admit though, that, for given k, eventually in n, $\xi \mapsto \mathrm{E}_{(\xi,\beta)}[\varepsilon_n^*(\mathrm{kMedMad}(k))]$ is decreasing s.t.

$$\lim_{\xi\to\infty} \mathrm{E}_{(\xi,\beta)}[\varepsilon_n^*(\mathrm{kMedMad}(k))] = 0 \tag{2.84}$$

At the same time, eventually in n, $\xi \mapsto \mathrm{E}_{(\xi,\beta)}[\varepsilon_n^*(\mathrm{PE}*)]$ is increasing with

$$\lim_{\xi\to\infty} \mathrm{E}_{(\xi,\beta)}[\varepsilon_n^*(\mathrm{PE}*)] = 1/4 \tag{2.85}$$

In particular, for $k = 10$, for $\xi \geq 4.964$, PE* has a better EFSBP / ABP, in this case $\bar{\varepsilon}^*(\mathrm{PE}*) \geq 19.0\%$. On the other hand, eventually in n, the EFSBP of kMedMad for the optimal $k = k(\xi)$ never drops below 32.1% for $\xi \in (0, 10]$ and below 25% for $\xi \in (0, 437]$, and achieves 39.9% for $\xi = 7.20$.

2.7.2 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 Tables 8 and 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. More specifically, 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 $n = 100$. As asymptotically, Hybr will coincide with kMedMad, $k = 10$, its asymptotic properties IF, $\mathrm{tr\,asVar}$, asBias are those of kMedMad, $k = 10$. In each case (default or failure) we have equivariance.

2.8 Maximally bias-robust Estimator: MBRE

If we only look at bias and want to obtain the procedure achieving minimax bias on the 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 $\omega_c^{\min}$ (see below).

Note however that contrary to Dupuis (1998), Dupuis and Field (1998) who use the Euclidean norm in the weighting function, we use the non-Euclidean norm n_β from (1.10) to achieve the discussed invariance.

Its optimality is determined solely by its $\mathrm{IF}\ \bar\psi$ 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

$$\mathrm{MBRE} = \theta_n^{(0)} + \frac{1}{n}\sum_{i=1}^{n} \bar\psi_{\theta_n^{(0)}}(X_i) \tag{2.86}$$

Minimizing asBias among all ALEs, we may read off the general solution from (Rieder, 1994, Thm. 5.5.1(b)), the minimal gross error sensitivity is given by

$$\omega_c^{\min} = \max\left\{ \mathrm{tr}\, d_\beta^{-1} A d_\beta^{-1} \Big/ \mathrm{E}\, n_\beta(A\Lambda - a), \qquad a \in \mathbb{R}^2, 0 \neq A \in \mathbb{R}^{2\times 2} \right\} \tag{2.87}$$

and $\mathrm{IF}\ \bar\psi$,

$$\bar\psi = \omega_c^{\min}(A\Lambda - a)/n_\beta(A\Lambda - a) \tag{2.88}$$

(the event $\{A\Lambda - a = 0\}$ carries probability 0). Apparently, (2.88) 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 $\omega_c^{\min}$, though. Corresponding algorithms to determine A, a, and $\omega_c^{\min}$ are implemented to R within the `ROptEst` package Kohl and Ruckdeschel (2009) available on CRAN.

Remark 2.5. Although the algorithms are implemented for general L_2-differentiable models there, particular algorithms and techniques are needed for the computation of the expectations under GPD (with its heavy tails)—essentially we integrate after a logarithmic substitution.

In our model, we obtain

$$\begin{aligned} A = A_{\mathrm{MBRE}} &= \begin{pmatrix} 1.000 & -0.183 \\ -0.183 & 0.224 \end{pmatrix}, \qquad a = a_{\mathrm{MBRE}} = (-0.179, 0.000), \\ \omega_c^{\min} &= 3.665 \end{aligned} \tag{2.89}$$

The use of norm n_β enforces invariance/equivariance,

$$\bar\psi_{(\xi,\beta)}(x) = d_\beta \bar\psi_{(\xi,1)}(x/\beta) \tag{2.90}$$

or, suppressing subscript MBRE, with

$$Y_{(\xi,\beta)} = A_{(\xi,\beta)}\Lambda_{(\xi,\beta)}(x/\beta) - a_{(\xi,\beta)} \tag{2.91}$$

$$\begin{aligned} A_{(\xi,\beta)} &= d_\beta\, A_{(\xi,1)}\, d_\beta, \qquad a_{(\xi,\beta)} = d_\beta\, a_{(\xi,1)}, \qquad n_\beta(Y_{(\xi,\beta)}) = n_1(Y_{(\xi,1)}), \\ \text{and } \omega_c^{\min}(\xi,\beta) &= \omega_c^{\min}(\xi,1) \end{aligned} \tag{2.92}$$

2.9 Estimator minimizing maximal MSE: OMSE

To get an estimator minimizing maximal MSE on neighborhoods (OMSE), we proceed similarly as in the case of the MBRE: We only 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

$$\mathrm{OMSE} = \theta_n^{(0)} + \frac{1}{n}\sum_{i=1}^{n} \hat\psi_{\theta_n^{(0)}}(X_i) \tag{2.93}$$

Again as starting estimator $\theta_n^{(0)}$ we use Hybr. In the general L_2 differentiable setting, the form of $\hat\psi$ may be read off from (Rieder, 1994, Thm. 5.5.7):

$$\hat\psi = Y \min\{1, b/n_\beta(Y)\}, \quad Y = A\Lambda - a \tag{2.94}$$

where $A \in \mathbb{R}^{2\times 2}$ and $a \in \mathbb{R}^2$ are such that $\hat\psi$ is an IF, i.e., (1.18) holds, and b is such that

$$r^2 b = \mathrm{E}(|Y| - b)_+ \tag{2.95}$$

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 `ROptEst` package available on CRAN. In our model, we obtain

$$\begin{aligned} A = A_{\mathrm{OMSE}} &= \begin{pmatrix} 10.258 & -2.894 \\ -2.894 & 3.869 \end{pmatrix}, \qquad a = a_{\mathrm{OMSE}} = (-1.076, 0.121), \\ b_{\mathrm{OMSE}} &= 4.401 \end{aligned} \tag{2.96}$$

As for MBRE, the use of norm n_β enforces invariance/equivariance,

$$\hat\psi_{(\xi,\beta)}(x) = d_\beta \hat\psi_{(\xi,1)}(x/\beta) \tag{2.97}$$

or again, (without the expression $\omega_c^{\min}$ and after suppressing OMSE), corresponding equations (2.91) and (2.92) together with

$$b_{(\xi,\beta)} = b_{(\xi,1)} \tag{2.98}$$

Remark 2.6. In fact, compare (Rieder, 1994, Thm. 5.5.7), OMSE also solves the "Lemma 5 problem" for bias bound its own GES, hence it is a particular OBRE in the terminology of Dupuis (1998), Dupuis and Field (1998).

The cited references, though, do not pursue the goal to find the MSE-optimal bias bound, and in this sense our OMSE will in general beat their OBRE (w.r.t. MSE at our radius r, of course).

On the other hand, for given bias bound b, (2.95) may be divided by b, and hence gives a radius $r(b)$ for which a given OBRE is MSE-optimal; in this sense, bias bound b and radius r are equivalent parametrizations of the degree of robustness required for the solution.

Computational Aspects Due to the lack of (complete) invariance, solving for equations (2.94) and (2.95) 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 dimension due to scale invariance, i.e.; we only would need to know the influence curves for "all" values $\xi > 0$. To speed up things, especially for our simulation study, we thus have used the following approximative approach, already realized in M. Kohl's R package `RobLox` for the Gaussian one-dimensional location and scale model[4], Kohl (2009):

Algorithm 2.7 (Lagrange multipliers by interpolation). In an offline phase, for a grid of size M, say $M = 200$, values of ξ, giving parameter values $\theta_i = (\xi_i, 1)$ and—in our case—to given radius $r = 0.5$, we determine the optimal IF's $\hat\psi_{\theta_i}$, solving equations (2.94) and (2.95) for each θ_i; for each of these, we 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 $A^\natural$, $a^\natural$, and $b^\natural$ by simple inter-/extrapolation for the stored grid values A_i, a_i, b_i. This gives us $Y^\natural = A^\natural \Lambda_{\theta'} - a^\natural$, and $w^\natural = \min\big(1, b^\natural / n_\beta(Y^\natural)\big)$. So far, $Y^\natural w^\natural$ would not make for an IF at θ'; thus, similarly to (Rieder, 1994, Rem. 5.5.2), we generate an approximating IF $\psi^\sharp$ by defining

$$z^\sharp = \mathrm{E}_{\theta'}[\Lambda_{\theta'} w^\natural] / \mathrm{E}_{\theta'}[w^\natural], \qquad A^\sharp = \Big\{ \mathrm{E}_{\theta'}[(\Lambda_{\theta'} - z^\sharp)(\Lambda_{\theta'} - z^\sharp)^\tau w^\natural] \Big\}^{-1}, \tag{2.99}$$

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

The solution produced in this algorithm will in general not (yet) solve (2.94) and (2.95), though, i.e. $A^\natural \neq A^\sharp$, $a^\natural \neq a^\sharp$, and equality will not hold in (2.95), but if the grid is dense enough, due to the smoothness of our model, we will have approximate equality in all these equations. This smoothness can be seen in Figure 4.

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.

[4] Due to the affine equivariance of the estimators in the location and scale setting, interpolation in package `RobLox` is done only for varying radius r.

The main advantage of Algorithm 2.7 is speed: While solving equations (2.94) and (2.95) will take about 15sec per ξ-value (hence per estimator evaluation), with the interpolation technique we can now produce 1000 evaluations in 120sec[5] (where most of the time is now consumed by producing the starting estimate, Hybr).

It also turns out that, up to accuracy 10^{-3}, we may even skip the recentering and restandardizing for IF, hence skipping five one-dimensional integrations, and instead directly work with $Y^{\natural}w^{\natural}$. This gives an extra performance gain of factor $5-10$, so all in all we may achieve a speed-up of around factor 1000. In our simulations study, however, we observed that for small samples, i.e., $n=40$, without the recentering and restandardizing for IF, we can only achieve about 90% efficiency.

Remark 2.8. Algorithm 2.7 applies to all ALEs which enjoy the partial (β-) invariance used here, and which involve solving for corresponding equations / finding minima, and where we may employ estimators constructed as one-step-estimators; this holds in particular for MBRE where we may allow for different pairs (A, a) in the nominator and denominator of the optimal term in (2.88).

Similar constructions could be used to store solutions for the implicit equations for MMed and (k)MedMad on a grid of ξ-values, and then for evaluation of the estimator use again inter-/extrapolation; we have done so for MMed, but not (yet) for (k)MedMad (and Hybr), where timings as for MMed should be in reach by this technique.

3 Synopsis of the Theoretical Properties

In a condensed form, in Table 6, we summarize our findings so far, evaluating criteria finite sample breakdown point FSBP (where possible), $\mathrm{asBias} = r\,\mathrm{GES}$, trace of the asymptotic variance asVar, and maximal asymptotic MSE on the neighborhood asMSE. $\mathrm{tr}\,\mathrm{asVar}$ and asMSE are evaluated on a quadratic scale, asBias on a linear scale; 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).

For FSBP, we evaluate terms at sample size $n = 1000$, which is relevant for MLE, SMLE (due to shrinking skipping rate of $r' = 0.7$, or $\alpha_n = 2.2\%$).
We also determine efficiencies in the ideal model and under contamination of radius 0.5 denoted by $\mathrm{eff.id}$ and $\mathrm{eff.re}$, respectively, as well as the respective ranks. In addition, for the situation where r is unknown, we also compute the least favorable efficiency of each (fixed) estimator (i.e.; we still use $r = 0.5$ for OMSE, although this is probably false) w.r.t. the most efficient procedure knowing the radius, denoted by $\mathrm{eff.ru}$ and again report the respective ranks (for this notion, cf. Rieder et al. (2008)). These efficiencies may be read as the relative amount of observations, the optimal procedure (MLE in the ideal setting, OMSE for $r = 0.5$ under contamination, and OMSE

[5] Times measured on a recent Dual Core Laptop.

for least favorable actual radius) would need to achieve the same accuracy as the estimator under consideration. Paralleling Kohl (2005, Lemma 2.2.3), we see that for all considered estimators S_n

$$\mathrm{eff.ru}(S_n) = \min\left(\mathrm{eff.id}(S_n), \mathrm{GES}^2(\mathrm{MBRE})/\mathrm{GES}^2(S_n)\right) \tag{3.1}$$

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

Finally, we document the ranges of least favorable x-values $x_{\mathrm{l.f.}}$, at which the considered IFs take their maximum in n_β-norm. Infinitesimally, these are the most vulnerable points of the corresponding estimators, as contamination placing mass therein will render bias maximal. The value ∞ appearing here is to be taken as a limit; in all considered situations, a value of 10^{10} will suffice to produce a (nearly) maximal bias. On the other hand, the Pickands-type estimators PE, PE*, and PicM, as well as MMed and the original MedMad estimator are most harmfully contaminated by placing extra mass at smallish values of, say, about $x = 1.5$ (for $\beta = 1$).

The classical PE estimator as well as MedMad are improved in all categories by their generalizations PE* and kMedMad (i.e.; with $k = 10$), so should be replaced by them. Among the explicit estimators, both PE* and PicM can achieve convincing values of asMSE (with slight advantages for PicM in the ideal model)—although both at the cost of a breakdown point of only $6\% - 7\%$. 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. As shown in Lemma 2.1, B_n is of exact unstandardized order $\mathrm{O}(\log(n)/\sqrt{n})$, hence consequently, asBias and asMSE should both be ∞, and the efficiencies in ideal and contaminated situation would both be 0. At sample size $n = 1000$, though, asBias and asMSE are finite: According to approximation (6.2), B_n at $n = 1000$ is 0.17 (unstandardized), respectively, multiplied by $\sqrt{n}$, 5.38, while the entry of 3.75 in Table 6 is just $r \sup|\,\mathrm{IF}\,|$. and is at large due to an underestimation of ξ by 0.17.

As already noted, MLE achieves smallest $\mathrm{tr\,asVar}$, hence cannot be beaten in the ideal model, but at the price of a minimal FSBP and an infinite gross error sensitivity, so one (extremely large) observation at any sample size suffices to render MSE arbitrarily large.
Although not explicit, 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 parameter values. If we believe in our conjectured FSBP of 35%, MDE shares this reliability with Hybr, but contrary to the former needs a reliable starting value for the optimization (which in fact can be given by Hybr). As to computation, it is quite fast though.
MBRE and OMSE are constructed as one-step estimators, so using a starting estimator with a high FSBP like Hybr, they inherit this property while, consistently to the theory, at the same time MBRE achieves lowest gross error sensitivity (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.

With respect to least favorable efficiency $\mathrm{eff.ru}$, OMSE for $r = 0.5$ is best among all 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 not so convincing.

estimator	asBias	tr asVar	asMSE	eff.id	rk.id	eff.re	rk.re	eff.ru	rk.ru	$x_{1.f.}$	$\bar{\varepsilon}^*_{1000}$
MLE	∞	6.29	∞	1.00	1	0.00	11	0.00	11	∞	0.00
PE	4.08	24.24	40.87	0.26	10	0.35	9	0.20	8	$[0.89; 2.34]$	0.06
PE*	3.59	18.23	31.08	0.34	9	0.45	7	0.26	6	$[1.41; 4.18]$	0.07
PicM	3.78	17.35	31.64	0.36	8	0.45	8	0.24	7	$[1.41; 2.34]$	0.06
MMed	2.62	17.45	24.32	0.36	7	0.58	6	0.32	5	$[0.00; 0.34] \cup [0.90; 2.54]$	$0.25^?$
MedMad	4.32	29.55	48.22	0.21	11	0.29	10	0.18	9	$[0.00; 0.18] \cup [0.90; 1.60]$	0.19
kMedMad	2.19	12.80	17.60	0.49	5	0.80	4	0.49	3	$[0.54; 0.89] \cup [4.42; \infty)$	0.31
SMLE	3.75	7.03	21.08	0.90	2	0.67	5	0.03	10	$[20.67; \infty)$	0.02
MDE	2.45	9.76	15.74	0.64	4	0.90	2	0.56	2	$\{0, \infty\}$	$0.35^?$
MBRE	1.84	13.44	16.80	0.47	6	0.84	3	0.47	4	$[0.00; \infty)$	0.35*
OMSE	2.20	9.73	14.13	0.64	3	1.00	1	0.68	1	$[0.00; 0.07] \cup [5.92; \infty)$	0.35*

Table 6 Comparison of the asymptotic robustness properties of the estimators
*: inherited from starting estimator Hybr; $^?$: conjectured.

In Figures 5(a) and (b), we display the influence curves (ICs) of the considered estimator. All considered ICs ψ_θ share the invariance property that $\psi_{(\xi,\beta)}(x) = d_\beta \psi_{(\xi,1)}(x/\beta)$. For completeness, we also include MME, although it is not available for our reference parameter value $(\xi = 0.7, \beta = 1)$; to this end, we use $(\xi = 0.2, \beta = 1)$ in this case; as is clear from (2.16), the ICs here are linear combinations of a linear function and a parabola, hence again parabolas and thus—compared to MLE—drastically unbounded.

All ICs to robust estimators for scale are redescenders, while those for shape are bounded and strictly positive for large enough x. All curves displayed in Figure 5(b) only take finitely many values—3 in case of PE and PE*, 4 for MMed, MedMad and kMedmad, and 31 for PicM—which makes integration quite easy.

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 optimal one (for the ideal or contaminated situation, respectively). 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.

In order to show that the choice of $\xi = 0.7$ gives "typical" results concerning the obtainable efficiencies, i.e. that the conclusions we just have drawn as to the ranking of the procedures remain valid for other parameter values, we have produced Figure 6 which also considers our estimators at $r = 0.25$ and $r = 1.0$ (without changing OMSE to this new r, though). Note that due to the scale invariance we do not need to consider other parameter values for β. From this figure we may in particular read off the minimal value for the efficiencies as extracted in Table 7.

estimator	MLE	PE	PE*	PicM	MMed	MedMad	kMedMad	SMLE	MDE	MBRE	OMSE
$\min_\xi \mathrm{eff.id}$	1.00	0.16	0.26	0.24	0.07	0.14	0.40	0.00	0.45	0.41	0.58
$\min_\xi \mathrm{eff.re}$	0.00	0.24	0.38	0.33	0.12	0.23	0.78	0.00	0.69	0.78	1.00
$\min_\xi \mathrm{eff.ru}$	0.00	0.15	0.22	0.18	0.07	0.14	0.40	0.00	0.43	0.41	0.58

Table 7 Minimal efficiencies for $\xi \in [0, 2]$ in the ideal model and for contamination of known and unknown radius

4 Simulation Study

In order to assess the finite sample properties of our estimators, we have done an extensive simulation study.

4.1 Setup

For sample sizes $n = 40, 100, 1000$, we simulate data from both the ideal GPD with parameter values $\mu = 0$, $\xi = 0.7$, $\beta = 1$. As estimators, we consider the same estimators as in the preceding section, and evaluate them at $M = 10000$ runs in the respective situation (ideal/contaminated and sample size $n = 40, 100, 1000$). In addition to these, we compute OMSE and MBRE in two variants of Algorithm 2.7, i.e., with IF–correction by recentering and restandardization (suffix w.c.) or without this correction, (suffix n.c.).

The contaminated data stems from the (shrinking) Gross Error Model (1.12), (1.13) with starting radius $r = 0.5$. For sample size n, this amounts to actual contamination sizes of $r_{n=40} = 7.9\%$, $r_{100} = 5\%$, and $r_{1000} = 1.6\%$. As contaminating data distribution, we use $G_{n,i} = \mathrm{Dirac}(10^{10})$, except for estimators PE, PE*, PicM, MMed and MedMad, where we use $G'_{n,i} = \mathrm{unif}(1.42, 1.59)$ in accordance with $x_{\mathrm{l.f.}}$ from Table 6.

For the resulting estimates, we compute empirical Bias_ξ, Bias_β, $n_\beta(\mathrm{Bias})$, Var, MSE. We also document the frequency of failures, and the computation time. Based on empirical risk, i.e.; (standardized) MSE, we determine efficiencies w.r.t. the corresponding optimal risk.

For MMed and kMedMad it turns out that, for maximal MSE it is preferable to 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 the NA's we use $G'_{n,i}$.

4.2 Results

Due to space restrictions, we only present a subset of our tables and plots.

In Tables 8, 9, and 10, we summarize the results for sample sizes $n = 40$, 100, and 1000, respectively. The first two columns show the sign of the bias in coordinate β and ξ, s_β, and s_ξ respectively; for values larger than 10 in absolute value we write "$--$" or "$++$", respectively, while for values not significantly deviating from 0 (at empirical significance 95%) we write ".". Values for $|\mathrm{Bias}|$, for variance, and for MSE (standardized by $\sqrt{n}$ and n, 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 solve for corresponding zeros, which due to the use of Hybr as starting estimators is then propagated to MLE, SMLE, MDE, MBRE, and OMSE. Column "time" gives the computation time in seconds on a recent dual core processor for the $M = 10000$ evaluations of the estimator at sample size n, aggregating time for ideal and contaminated situation. These timings, of course, are subject to future advances in both hardware and OS, but the relative timings should remain relevant. 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 optimal estimator is printed in bold face.

The simulation study confirms our findings of section 3; figures, at least for $n = 1000$, are—at large—close to the ones of Table 6. This holds in particular for the ideal situation, and for the efficiencies, where in the latter case we obtain reasonable approximations already for $n = 100$—at the exception of SMLE and the PE-variants.

Remark 4.1. This is consistent to higher order asymptotics for the MSE as developed in Ruckdeschel (2010b): In both the ideal situation, and, uniformly, on corresponding shrinking gross error neighborhoods, MSE allows asymptotic expansions of form

$$n\mathrm{MSE} = A_0 + rA_1 n^{-1/2} + A_2 n^{-1} + \mathrm{o}(n^{-1}) \tag{4.1}$$

where term A_0 is $\mathrm{tr}\,\mathrm{asVar}$ in the ideal model and asMSE in the contaminated situation, and A_1, A_2 are terms depending on of functions of type $t \mapsto \mathrm{E}_\theta[\psi_t^k \eta_t^l]$, $k, l \in \mathbb{N}$, and their respective derivatives w.r.t. the parameter as well as $t \mapsto \sup|\psi_t|$, $t \mapsto \sup|\eta_t|$, where η is the IF of the starting estimator. Hence, in particular the first correction term in the ideal situation ($r = 0$) is of order $\mathrm{O}(1/n)$, while in the contaminated situation, it is of order $\mathrm{O}(1/\sqrt{n})$.

Grossly speaking, 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 $\mathrm{OMSE}_{\mathrm{w.c.}}$ at significance 95% is the best estimator among the considered ones as to MSE, although, especially for small sample sizes, MDE, $\mathrm{MBRE}_{\mathrm{w.c.}}$, 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 cannot be said for MMed, MedMad, and kMedMad, which suffer from up to 33% failure rate at this sample size under contamination. So Hybr is a real improvement.

For small sample sizes IF-correction pays off significantly in terms of MSE—variant $\mathrm{OMSE}_{\mathrm{n.c.}}$ without this correction at sample size 40 looses 11% in efficiency w.r.t. $\mathrm{OMSE}_{\mathrm{w.c.}}$ in the contaminated model.

Curiously, at $n = 1000$, the bias B_n of SMLE present in the ideal situation is decreased by contamination, to the effect that here our asymptotic value $\mathrm{asMSE}_{\mathrm{re}}$ (21.39) is astonishingly accurate as approximation for $\mathrm{MSE}_{\mathrm{re}}$ (again 21.39). Still, according to the empirical values, one would not recommend SMLE (at least not without a bias correction).

Among the optimal procedures, there is a distinction between cases $n = 40, 100$ and $n = 1000$. In the first case, evaluating the corresponding integrals needed for the correction for IF in $\mathrm{OMSE}_{w.c.}$ are most expensive, taking about twice the time of the competitors. At sample size 40 a compromise would use $\mathrm{MBRE}_{w.c.}$ with approximately the same efficiency (under contamination) and needing about half the time, while at sample size 100, one could also use $\mathrm{OMSE}_{n.c.}$ which only takes $1/10$ of the time while loosing only 4% in efficiency (under contamination). At sample size 1000, $\mathrm{OMSE}_{n.c.}$ even beats $\mathrm{OMSE}_{w.c.}$ slightly, consuming only less than $1/5$ of the computation time. At this sample size, though, this effect is dominated by the time needed for kMedMad: Used as starting estimator for $\mathrm{OMSE}_{n.c.}$, we spend roughly 90% of the time computing Hybr. Still, this is not this bad: on average we need about $.07$ seconds for computing one estimator at $n = 1000$; for comparison: this time decreases to $.016$ seconds for $n = 100$ and $.014$ seconds for $n = 40$.

The results for sample sizes $n = 40, 100, 1000$ are graphically displayed in boxplots in Figures 7(a)–9(b), respectively. In Figure 7(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 MedMad, PE, and MBRE.nc are somewhat less precise; under contamination, as illustrated in Figure 7(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. The transition to $n = 100$, and even more so to $n = 1000$, most strikingly increases accuracy for all estimators, as we would expect. The bias of SMLE in the ideal model gets smaller, but remains visible, while differences between competitors OMSE.wc, MDE, kMedMad, and OMSE.nc, MBRE.[n/w]c, are harder to spot at the uniform scale, which is why we include enlarged Figures 10(a)–11(b); in these we see that under contamination, there is underestimation of ξ by the PE-type, MMed, and SMLE estimators resp. overestimation by the remaining estimators, while β is overestimated by the PE-type, MMed, MedMad, and (less so) SMLE, and underestimated by MLE, the remaining estimates for β stay largely unbiased even under contamination.

5 Conclusion

We have compared MLE, SMLE, MDE CvM, PE, PE*, PicM, MMed, MedMad, kMedMad, Hybr, MME, and the optimally robust MBRE and OMSE as estimators for scale and shape parameters ξ and β of the generalized Pareto distribution on the ideal and contaminated data in terms of local and global robustness properties.

Asymptotic theory and empirical simulations show that as to global reliability, Hybr, kMedMad, MDE, MBRE, and OMSE estimators can withstand relatively high rates of outliers as expressed by a high (E)FSBP of roughly $1/3$. Much less so, but still with considerably positive values of (E)FSBP, we have MedMad, and even less, PE*, PE, and PicM. SMLE in the variant without bias correction as used in this paper, but with shrinking skipping rate, and MLE, and MME all have minimal FSBP of $1/n$, hence should be avoided.

High failure rates for MMed, MedMad, and kMedMad at small sample sizes, in particular under contamination makes their use prohibitive.

Looking at the infinitesimal effects of one observation on the estimator, as expressed through the influence function, we see that, except for MLE and MME, 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 different asymptotic risks, as well as (simulated) finite sample risks: Overall, we can recommend estimator OMSE with Hybr as starting estimator; it has achieved best risk in the simulations, may be computed fast, is efficient (100%) for contamination of known radius and for parameter value $\xi \in [0, 2]$ never drops below 58% efficiency in the ideal model and for contamination of unknown radius.

To be fair, one has to say, that MBRE, and MDE come quite close to OMSE, for parameter value $\xi \in [0, 2]$ never falling below 78%, resp. 69% efficiency under contamination and similarly in the ideal model (MDE 45% and MBRE 41%) and under contamination of unknown radius (43% resp. 41%).

Among the (almost) explicit estimators, clearly kMedMad (resp. Hybr) stands out and comes closest to the aforementioned group—minimal efficiency $\xi \in [0, 2]$ not below 78% for contamination of known radius and 40% in the ideal setting resp. for contamination of unknown radius.

MedMad, and even more so, the Pickand variants PE, PE*, and PicM are also robust, but not really advisably due to their low breakdown points, and, additionally, due to their non-convincing efficiencies; the only reason for using PE, PE*, (and less so PicM) is their ease of computation, which should not be so decisive, though.

Still, they beat the popular SMLE without bias correction, which does provide some, but much too little protection against outliers.

Worst, of course, as to robustness aspects are MLE and MME, where the latter in addition has a limited application range.

6 Proofs

PROOF TO LEMMA 1.3: Assume without loss that $\sqrt{n}(\theta' - \theta) \to h$; then by L_2-differentiability together with (1.16)

$$\mathrm{E}_{\theta'}[\psi_\theta] = \mathrm{E}_\theta[\psi_\theta] + \mathrm{E}_\theta[\psi_\theta \Lambda_\theta^\tau]h/\sqrt{n} + \mathrm{o}(n^{-1/2}) \tag{6.1}$$

so under $P_{\theta'}^n$, for R_n' the remainder of (1.15) for $\theta' = \theta$ we obtain

$$\begin{aligned}\sqrt{n}\,(S_n - \theta) - h - R_n &\overset{(1.15)}{=} \frac{1}{\sqrt{n}}\sum_{i=1}^n (\psi_\theta(X_i) - \mathrm{E}_{\theta'}[\psi_\theta]) = \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n (\psi_\theta(X_i) - \mathrm{E}_\theta[\psi_\theta]) - \mathrm{E}_\theta[\psi_\theta \Lambda_\theta^\tau]h + \mathrm{o}(1) = \\ &= \sqrt{n}\,(S_n - \theta) - \mathrm{E}_\theta[\psi_\theta \Lambda_\theta^\tau]h - R_n' + \mathrm{o}(1)\end{aligned}$$

Again by L_2-differentiability, all $P_{\theta'}^n$ for $|\theta' - \theta| = \mathrm{O}(n^{-1/2})$ are mutually contiguous, so $\sqrt{n}\,R_n$, $\sqrt{n}\,R_n'$ both converge to 0 stochastically under both $P_{\theta'}^n$ and P_θ^n, so necessarily (1.17) holds. This also shows the second assertion. ////

PROOF TO PROPOSITION 1.8: We start with the fact that for $X_i \overset{\text{i.i.d.}}{\sim} F$ with Lebesgue density f, the joint c.d.f. of the order statistics $X_{[i_1:n]}$, $X_{[i_2:n]}$ for $1 \le i_1 < i_2 \le n$ for $s \le t$ can be written as

$$G(s,t) \;=\; n\int_{-\infty}^{s} f(s)\tbinom{n-1}{i_1-1}F(s)^{i_1-1}\sum_{k_2=i_2-i_1}^{n-i_1}\tbinom{n-i_1}{k_2}\big(F(t)-F(s)\big)^{k_2}\bar F(t)^{n-i_1-k_2}\,ds$$

Hence

$$\begin{aligned}P(\hat N_n' \ge l) &= P(X_{[(n/2+l+1):n]} \le (k+1)X_{[(n/2+1):n]}) = \\ &= n\int_0^1 \tbinom{n-1}{n/2}u^{n/2}\sum_{k_2=l}^{n/2-1}\tbinom{n/2-1}{k_2}\big(F(q_k)-u\big)^{k_2}\bar F(q_k)^{n/2-1-k_2}\,du\end{aligned}$$

and (1.28) follows by taking differences. Cases (1.27) and (1.29) follow similarly. ////

PROOF TO LEMMA 2.1: We first note that $\alpha_0 < x_0$, the positive zero of $x \mapsto \log(1-x) + x + x^2$ ($x_0 \doteq 0.6837$). By the asymptotic linearity of the MLE, if we use a suitable (uniformly integrable) initialization, the bias of the SMLE has the asymptotic representation

$$\begin{aligned}B_n &= n_\beta(\mathrm{E}(\mathrm{SMLE}) - \theta) = n_\beta\Big(\frac{1}{n}\sum_{k=1}^{\alpha_n n}\mathrm{E}\,\mathrm{IF}_{(\xi,\beta)}(z(X_{(n+1-k:n)})); \mathrm{MLE}, F)\Big) = \\ &= \Big((\frac{1}{n}|\sum_{k=1}^{\alpha_n n}\mathrm{E}\,\tilde\psi_\xi(V_{(k:n)})|^2 + (\frac{1}{n}|\sum_{k=1}^{\alpha_n n}\mathrm{E}\,\tilde\psi_\beta(V_{(k:n)})|^2/\beta^2\Big)^{1/2}\end{aligned} \tag{6.2}$$

for $X_{(k:n)}$, $V_{(k:n)}$ the respective kth order statistic. Using (2.5), we see that for v ranging in $(0,1)$, the coordinates of the IF of MLE may each be written as $a\log(v)+f(v)$, $a\neq 0$, and f bounded on this range. Hence the dominating term is $\log(v)$, so we have to check the behavior of $|\operatorname{E}\log(B_{k,n})|$ for $B_{k,n}\sim \operatorname{Beta}(k,n-k+1)$, $k=1,\ldots,\alpha_n n$. To this end, note that by the power series expansion of $\log(1-x)$, for any $L>0$ and any $x\in(0,1]$, $-\log(x)\geq\sum_{l=1}^{L}(1-x)^l/l$, while for $0\leq x<x_0$, $\log(1-x)\geq -x-x^2$. We further observe that (for $n>k$), as $1-B_{k,n}\sim\operatorname{Beta}(n-k+1,k)$, $\operatorname{E}(1-B_{k,n})^l=\prod_{j=1}^{l}\frac{n+j-k}{n+j}$, and that for any decreasing suitably integrable function $f(x)$ with antiderivative $F(x)$, $\sum_{j=1}^{n}f(j)\leq\int_0^n f(x)\,dx=F(n)-F(0)$. Hence, using $1-x\leq e^{-x}$ for $x\in\mathbb{R}$ we obtain

$$|\operatorname{E}\log(B_{k,n})|\geq\sum_{l=1}^{L}\operatorname{E}(1-B_{k,n})^l/l\geq\sum_{l=1}^{L}\frac{1}{l}\prod_{j=1}^{l}\tfrac{n+j-k}{n+j}=$$

$$\begin{aligned}
&= \sum_{l=1}^{L}\tfrac{1}{l}\exp\big(\sum_{j=1}^{l}\log(1-\tfrac{k}{n+j})\big)\geq\sum_{l=1}^{L}\tfrac{1}{l}\exp\big(-\sum_{j=1}^{l}\tfrac{k}{n+j}+\tfrac{k^2}{(n+j)^2}\big)\geq\\
&\geq \sum_{l=1}^{L}\tfrac{1}{l}\exp\big(-k\log(\tfrac{n+l}{n})-\tfrac{k^2l}{(n+l)n}\big)=\sum_{l=1}^{L}\tfrac{1}{l}(1-\tfrac{l}{n+l})^k\exp(-\tfrac{k^2l}{(n+l)n})\geq\\
&\geq \sum_{l=1}^{L}\tfrac{1}{l}(1-\tfrac{L}{n+L})^k\exp(-\tfrac{k^2L}{(n+L)n})\geq\log(L)(1-\tfrac{L}{n+L})^k\exp(-\tfrac{k^2L}{(n+L)n})
\end{aligned}$$

Plugging in $L=\lceil\frac{1}{\alpha_n}\rceil$, we obtain, eventually in n,

$$E_{k,n}:=|\operatorname{E}\log(B_{k,n})|\geq-\log(\alpha_n)\exp(-1-\alpha_n)$$

On the other hand, for $d_{1,n}$ the densitiy of $\operatorname{Beta}(1,n)$, we split the integration range into $[0,1/n]$ and $[1/n,1]$ and obtain

$$0\;<\;\int_0^1-\log(x)\,d_{1,n}(x)\,dx\leq n(\log(n)+1)/n+\log(n)\leq 3\log(n)$$

if $n>2$. Now, for some $d_1,d_2\geq 0$

$$|\operatorname{E}\tilde\psi_\xi(B_{k,n})|=\tfrac{(\xi+1)^2}{\xi}E_{k,n}+d_1-\tfrac{\xi^2+3\xi+1}{\xi^2+\xi},\qquad|\operatorname{E}\tilde\psi_\beta(B_{k,n})|=\tfrac{(\xi+1)}{\xi}E_{k,n}+d_2-(3-\tfrac{1}{\xi})$$

Hence, as $\frac{\xi^2+3\xi+1}{\xi^2+\xi}<3+\xi^{-1}$, for $\liminf\alpha_n<\alpha_0$ we obtain, eventually in n

$$\begin{aligned}
0\;&\leq\;\tfrac{(\xi+1)\sqrt{(\xi+1)^2+\beta^{-2}}}{\xi}\alpha_n(-\log(\alpha_n/\alpha_0))\exp(-1-\alpha_n)\leq\\
&\leq\;\frac{1}{n}\sum_{k=1}^{\alpha_n n}\frac{\xi+1}{\xi}\sqrt{((\xi+1)^2+\beta^{-2})\big(E_{k,n}-3-1/\xi\big)^2}\leq\\
&\leq\;\Big(\{\frac{1}{n}\sum_{k=1}^{\alpha_n n}\operatorname{E}\tilde\psi_\xi(B_{k,n})\}^2+\{\frac{1}{n}\sum_{k=1}^{\alpha_n n}\operatorname{E}\tilde\psi_\beta(B_{k,n})\}^2/\beta^2\Big)^{1/2}=B_n
\end{aligned}$$

and $\liminf B_n > 0$ if $\liminf \alpha_n > 0$, resp. $\liminf n^\beta B_n > cn^\beta \alpha_n \log(n)$ if $\liminf n^\beta \alpha_n > 0$. On the other hand, eventually in n (as the other summand terms of $\tilde{\psi}$ are bounded in n)

$$B_n \leq 4\frac{(\xi+1)\sqrt{(\xi+1)^2+1/\beta^2}}{\xi^2}\alpha_n \log(n)$$

////

References

Basel Committee on Banking Supervision (2006) International Convergence of Capital Measurement and Capital Standards: A Revised Framework. http://www.bis.org/publ/bcbs128.pdf 1

Balkema, A. and de Haan, L. (1974) Residual life time at great age. Annals of Probability, 2: 792–804 1, 1.1

Baud, N., Frachot, A., Roncalli, T.: (2002) Internal data, external data and consortium data for operational risk measurement: How to pool data properly? SSRN: http://ssrn.com/abstract=1032529 Accessed 1 June 2002

Boudt, K., Caliskan, D. and Croux, C. (2010) Robust and Explicit Estimators for Weibull Parameters. Metrika, published online 28 July 2009, DOI 10.1007/s00184-009-0272-1 2.7, 2.7

Brazauskas, V. and Serfling, R. (2000) Robust Estimation of Tail Parameters for Two-Parameter Pareto and Exponential Models via Generalized Quantile Statistics. Extremes, 3(3): 231–249 1

Bickel, P. J. (1981) Quelques aspects de la statistique robuste. In Ecole d'Eté de Probabilités de Saint Flour IX 1979, (P.L. Hennequin, ed.): 1–72. Lecture Notes in Mathematics #876, Springer 1.1

Castillo, E. and Hadi, A. S. (1997) Fitting the Generalized Pareto Distribution to Data. Journal of American Statistical Association, 92(440): 1609–1620 1, 2.5

Christmann, A. and Van Messem, A. (2008) Bouligand Derivatives and Robustness of Support Vector Machines for Regression. Journal of Machine Learning Research, 9: 915–936 1.2

Cope, E. W., Mignola, G., Antonini, G., and Ugoccioni, R. (2009) Challenges and pitfalls in measuring operational risk from loss data. Journal of Operational Risk, 4(4), Winter 2009/10 1, 2.2

Davies, P. L. and Gather, U. (2005) Breakdown and groups (with discussion). Annals of Statistics, 33(3): 977–1035 1.2

Donoho, D. L. and Huber, P. J. (1983) A Festschrift for Erich L. Lehmann in honor of his sixty-fifth birthday. Wadsworth Statistics/Probability Series: 157–185 1.2

Donoho, D. L. and Liu, R. C. (1988) The "Automatic" Robustness of Minimum Distance Functionals. Annals of Statistics, 16(2): 552–586 2.4, 2.4

Donoho, D. L. and Liu, R. C. (1988) Pathologies of Some Minimum Distance Estimators. Annals of Statistics, 16(2): 587–608 2.2

Dupuis, D. J. (1998) Exceedances over high thresholds: A guide to threshold selection. Extremes, 1(3): 251–261 1, 2.8, 2.6

Dupuis, D. J. and Field, C. A.(1998) Robust estimation of extremes. Can. J. Stat, 26, 199–216 1, 2.8, 2.6

Dupuis, D. J. and Morgenthaler, S.(2002) Robust weighted likelihood estimators with an application to bivariate extreme value problems. Can. J. Stat., 30(1), 17–36 1, 2.2

Fernholz, L.T. (1979) von Mises Calculus for Statistical Functionals. Lecture Notes in Statistics #19, Springer 1.2

Field, C. and Smith, B. (1994) Robust Estimation—A Weighted Maximum Likelihood Estimation. International Review, 62(3): 405–424 2.2

Hadi, A.S. and Luceño, A. (1997) Maximum trimmed likelihood estimators: a unfied approach, examples, and algorithms. Computation Statitics & Data Analysis, 25:251–272 2.2

He, X. and Fung, W.K. (1999) Method of medians for life time data with Weibull models. Statistics in Medicine, 18:1993–2009 1

Hampel, F. R. (1986) Contributions to the theory of robust estimation. Ph.D. Thesis, University of California, Berkeley 1, 1.2, 1.2, 1.2

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986) Robust statistics. The approach based on influence functions. Wiley Series in Probability and Mathematical Statistics, Wiley 1, 1.1, 1.2, 1.5, 2.8

Hosking R. J. M. and Wallis T. J. (1987) Parameter and Quantile Estimation for the Generalized Pareto Distribution. Technometrics, 29(3) 1

Huber, P. J. (1981) Robust statistics, Wiley 2.2, 2.4

Huber-Carol, C. (1970) Étude asymptotique de tests robustes. Thèse de Doctorat, ETH Zürich 1.1

HORBENKO, N. AND RUCKDESCHEL, P. (2010) Non-smoothness of GPD Models *dd* 1.1

JUÁREZ, S. F. (2003) Robust and efficient estimation for the generalized Pareto distribution. *Unpublished Ph.D. dissertation, Statistical Science Department, Southern Methodist University*. Available at http://www.smu.edu/statistics/faculty/SergioDiss1.pdf 1

JUÁREZ, S. F., SCHUCANY, W. R. (2004) Robust and Efficient Estimation for the Generalized Pareto Distribution. Extremes, 7(3): 237–251 1

KOHL, M. (2005) Numerical Contributions to the Asymptotic Theory of Robustness. PhD Thesis, Universität Bayreuth. Also available under http://stamats.de/ThesisMKohl.pdf 3

KOHL, M. (2009) RobLox: Optimally robust influence curves and estimators for location and scale. R Package available in version 0.7 on CRAN, URL http://cran.r-project.org/ 2.9

Kohl, M., Rieder H., and Ruckdeschel, P. (2010) Infinitesimally Robust Estimation in General Smoothly Parametrized Models. To appear in Stat. Meth. & Appl. 1.2

KOHL, M. AND RUCKDESCHEL, P. (2009) ROptEst: Optimally robust estimation. R Package available in version 0.7 on CRAN, URL http://cran.r-project.org/. 1.3, 2.8

LINDE, M. (2007) Robuste Statistische Modellierung im Operational Risk Management - eine quantitative Analyse des operationellen Risikos im Bankensektor unter Verwendung robuster Parameterschätzverfahren. Diplomarbeit im Studiengang Wirtschaftsmathematik, Universität Duisburg-Essen 3

MÜLLER, C. H. AND NEYKOV, N. (2003) Breakdown points of trimmed likelihood estimators and related estimators in generalited linear models. J. Statist. Plann. Inference, 116: 503–519 2.2

NESLEHOVA, J., CHAVEZ-DEMOULIN, V., AND EMBRECHTS, P. (2006) Infinite Mean models and the LDA for operational risk. Journal of Operational Risk, 1(1): 3–25 1, 2.3

PENG, L. AND WELSCH, A. H. (2001) Robust Estimation of the Generalized Pareto Distribution. Extremes, 4(1): 53–65 1, 1, 2.6, 2.6

PICKANDS, J. (1975) Statistical Inference Using Extreme Order Statistics. Annals of Statistics, 3(1): 119–131 1, 1, 1.1, 2.5

R DEVELOPMENT CORE TEAM (2009) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org 1.3

RIBATET, M. (2009) POT: Generalized Pareto Distribution and Peaks Over Threshold. R Package available in version 1.1-0 on CRAN, URL http://cran.r-project.org/ 1.3

RIEDER, H. (1994) Robust Asymptotical Statistics. Springer Series in Statistics, Springer-Verlag. 1.1, 1.2, 1.2, 1.2, 1.2, 1.2, 2.1, 2.4, 2.4, 2.4, 2, 2.4, 2.5, 2.7, 2.8, 2.8, 2.9, 2.6, 2.7

RIEDER, H. (1978) A robust asymptotic testing model. Annals of Statistics, 6: 1080–1094 1.1

RIEDER, H., KOHL, M., AND RUCKDESCHEL, P. (2008) The Cost of not Knowing the Radius. Statistical Methods and Applications, 17(1): 13–40. 1.1, 3

RUCKDESCHEL, P. (2010a) Uniform integrability on Neighborhoods. *dd* 1.2, 2.1, 2.2

RUCKDESCHEL, P. (2010b) Uniform higher order asymptotics on neighborhoods for the MSE. *dd* 4.1

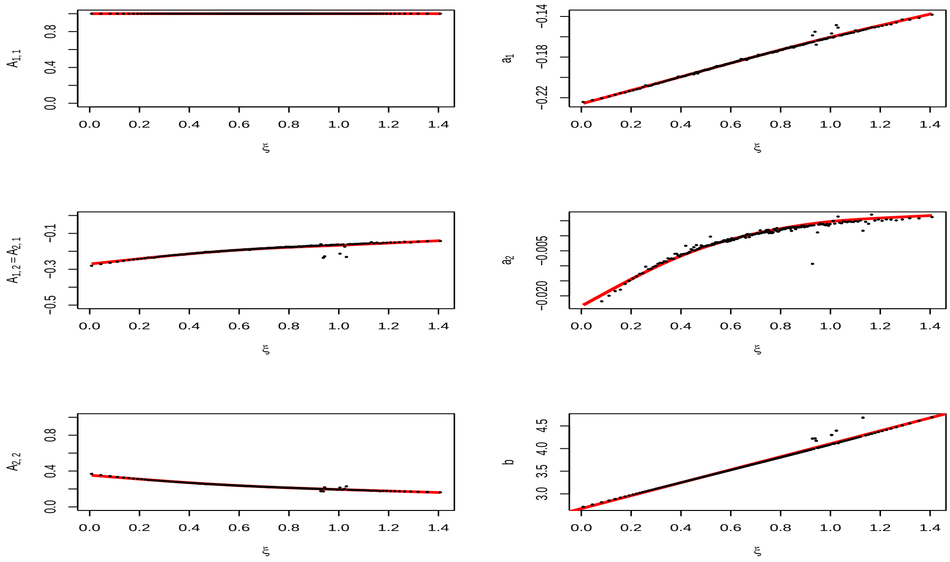
SERFLING, R., MAZUMDER, S. (2009) Exponential Probability Inequality and Convergence Results for the Median Absolute Deviation and Its Modifications. Statistics & Probability Letters, 79: 1767–1773 2.7

SMITH, L. R. (1987) Estimating tails of probability distributions. The Annals of Statistics, 15(3): 1174–1207 1

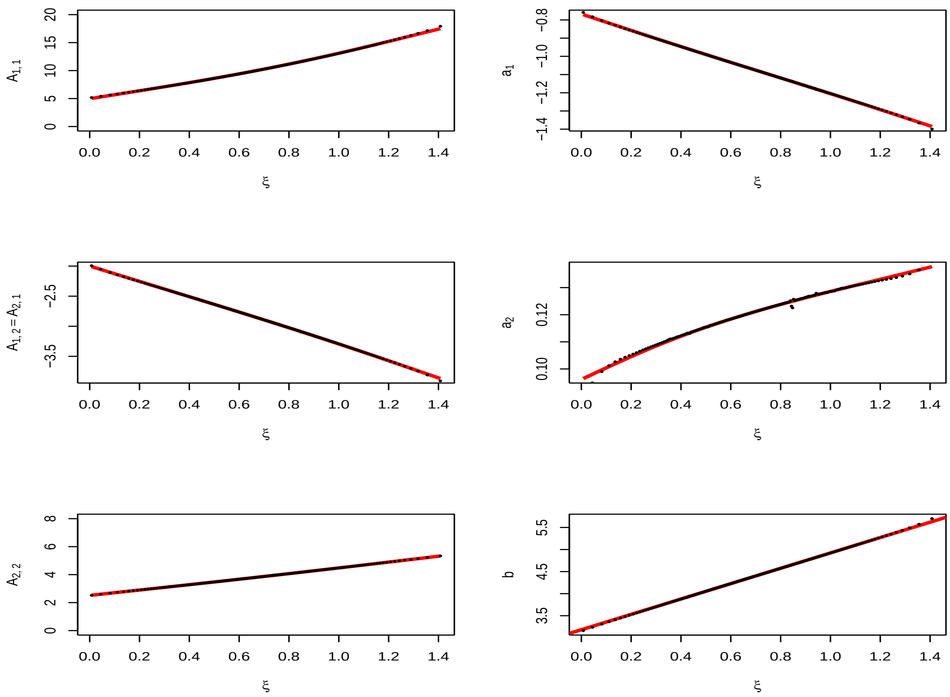
van der Vaart, A. W. (1998) Asymptotic statistics. Cambridge Univ. Press, Cambridge 1.1, 2.1

VANDEV, D. L., NEYKOV, N. M. (1998) About regression estimators with high breakdown point. Statistics, 32: 111–129 2.2

Witting, H. (1985) Mathematische Statistik I: Parametrische Verfahren bei festem Stichprobenumfang. B.G. Teubner, Stuttgart 1.1

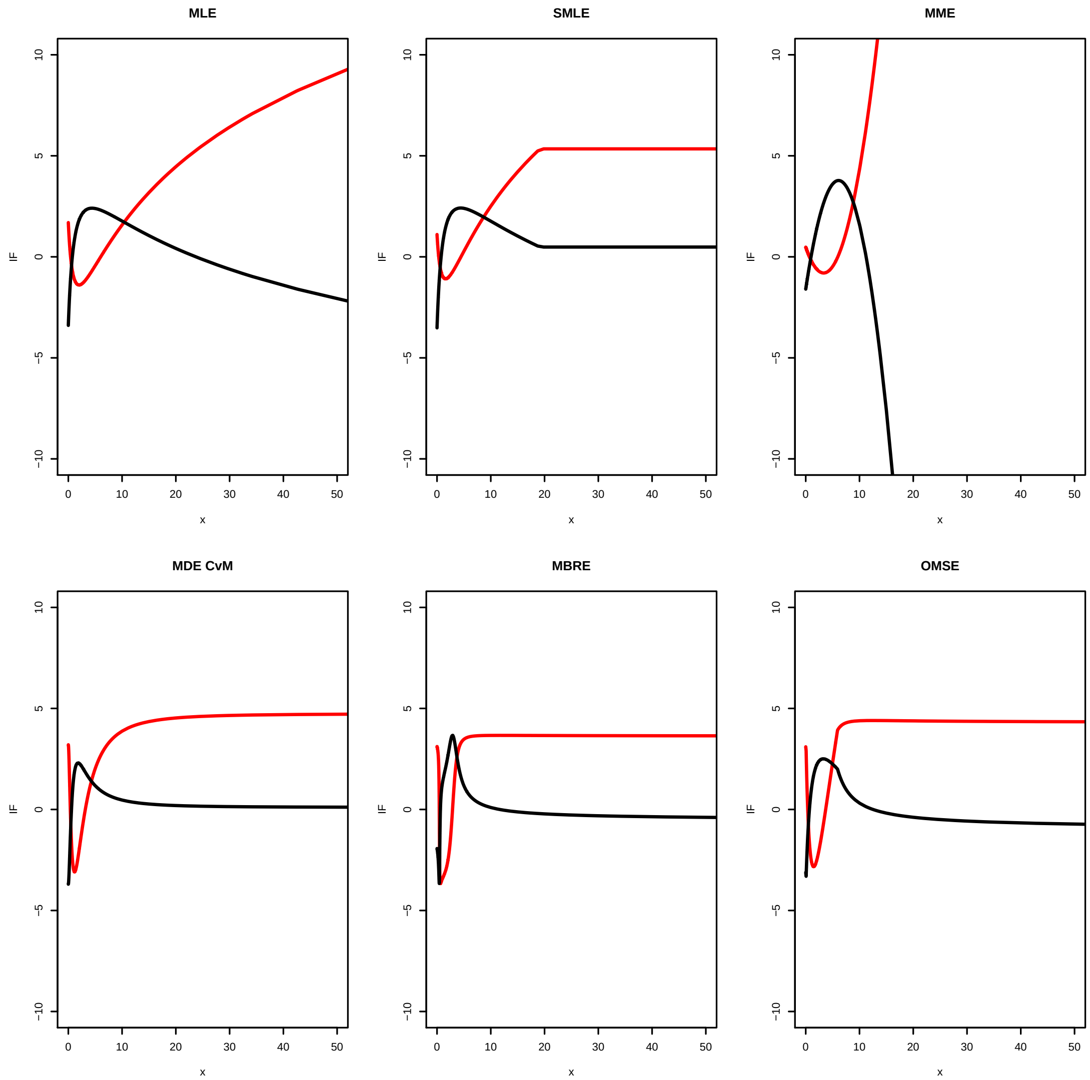


(a) Lagrange multipliers of MBRE at $\beta = 1$ (due to invariance) as functions in ξ

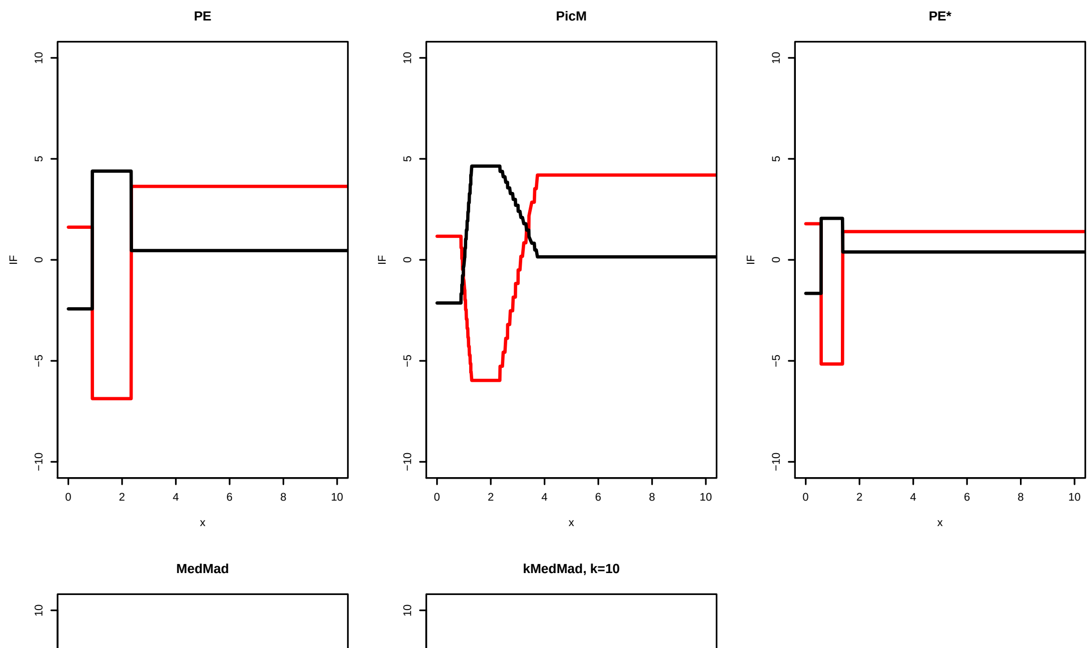


(b) Lagrange multipliers of OMSE at $\beta = 1$ (due to invariance) as functions in ξ

Figure 4: Lagrange multipliers of MBRE, OMSE at $\beta = 1$ (due to invariance) as functions in ξ: We see that we may easily interpolate between the grid points (depicted as circles) and lay corresponding smoothing splines (red curves; with R function `smooth.spline` for parameter `df=4`) through them.



(a) ICs of MLE, SMLE, MME, MDE CvM, MBRE, OMSE



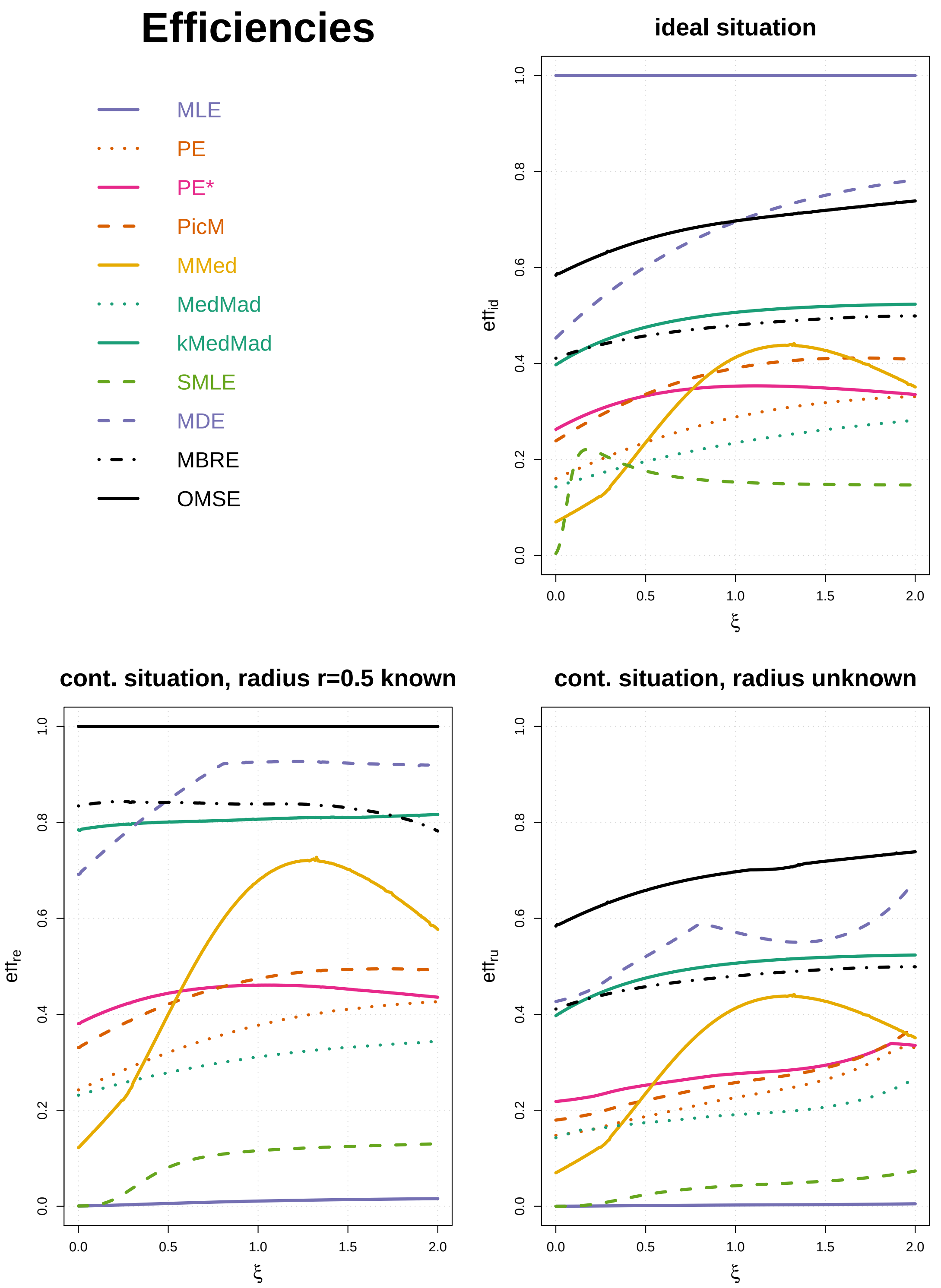


Figure 6 Efficiencies for varying shape
of MLE, SMLE (with $\approx 0.7 \cdot \sqrt{n}$ skipped value), MDE CvM, MBRE, OMSE, PE, MMed, kMedMad estimators for scale $\beta = 1$ and varying shape ξ.

ideal situation:

estimator	s_β	s_ξ	|Bias|	Var	MSE	eff	rank	NA	time
MLE	+	−	**0.55** ±0.05	**7.41** ±0.21	**7.72** ±0.21	**1.00**	**1**	**3.60**	**113**
PE	+	.	0.85 ±0.27	19.30 ±1.54	20.01 ±1.67	0.39	10	0.00	13
PE*	+	−	0.91 ±0.25	16.09 ±1.29	16.92 ±1.37	0.46	9	0.00	13
PicM	+	−	0.90 ±0.25	15.74 ±1.36	16.56 ±1.49	0.47	8	0.00	190
MMed	.	+	8.91 ±1.98	1.02 e5 ±2423.14	1.02 e5 ±2458.24	0.00	14	10.44	168
MedMad	−	+	1.32 ±0.10	24.77 ±1.30	26.52 ±1.39	0.29	12	21.42	150
kMedMad	+	−	0.47 ±0.07	11.55 ±0.30	11.78 ±0.29	0.66	4	8.08	197
Hybr	+	−	0.71 ±0.07	11.96 ±0.31	12.46 ±0.30	0.62	6	0.79	223
SMLE	+	−	4.70 ±0.06	9.49 ±0.30	31.62 ±0.47	0.24	13	0.79	75
MDE	+	−	0.40 ±0.06	10.56 ±0.27	10.72 ±0.25	0.72	3	0.79	384
MBRE.wc	+	−	0.49 ±0.08	15.68 ±0.46	15.92 ±0.44	0.48	7	0.79	302
OMSE.wc	+	−	0.26 ±0.06	9.62 ±0.23	9.68 ±0.22	0.80	2	0.79	600
MBRE.nc	+	−	0.80 ±0.09	19.39 ±0.53	20.03 ±0.52	0.39	11	0.79	38
OMSE.nc	+	−	0.95 ±0.07	11.36 ±0.34	12.25 ±0.33	0.63	5	0.79	41

contaminated situation:

estimator	s_β	s_ξ	|Bias|	Var	MSE	eff	rank	NA
MLE	+	++	394.12 ±22.92	1.37 e7 ±1.20 e6	1.52 e7 ±1.37 e6	0.00	14	3.61
PE	+	+	2.32 ±0.49	62.25 ±67.90	67.64 ±69.35	0.39	9	0.00
PE*	−	++	14.77 ±2.37	1456.35 ±256.97	1674.41 ±325.54	0.02	11	0.00
PicM	+	+	4.17 ±0.82	176.51 ±84.36	193.90 ±90.11	0.14	10	0.00
MMed	+	+	5.13 ±1.17	3563.54 ±1442.56	3589.87 ±1454.42	0.01	12	23.11
MedMad	+	+	1.01 ±0.10	23.58 ±1.46	24.61 ±1.44	0.89	6	33.08
kMedMad	+	+	2.32 ±0.09	18.82 ±0.49	24.21 ±0.67	0.91	5	19.10
Hybr	+	+	2.23 ±0.09	19.23 ±0.50	24.21 ±0.67	0.91	4	3.03
SMLE	+	+	7.44 ±3.10	2.51 e5 ±1.52 e5	2.52 e5 ±1.52 e5	0.00	13	3.61
MDE	+	+	2.64 ±0.08	16.19 ±0.43	23.15 ±0.59	0.95	2	3.61
MBRE.wc	−	+	1.77 ±0.09	20.06 ±0.59	23.19 ±0.63	0.95	3	3.03
OMSE.wc	−	+	**2.75** ±0.07	**14.39** ±0.42	**21.93** ±0.61	**1.00**	**1**	**3.03**
MBRE.nc	.	+	1.29 ±0.10	23.67 ±0.66	25.34 ±0.59	0.87	8	3.03
OMSE.nc	+	+	2.34 ±0.09	19.53 ±0.63	25.02 ±0.76	0.88	7	3.03

Table 8: Comparison of the empirical robustness properties of the estimators at $n = 40$

ideal situation:

estimator	s_β	s_ξ	|Bias|	Var	MSE	eff	rank	NA	time
MLE	**+**	**−**	**0.35 ±0.05**	**6.76 ±0.16**	**6.88 ±0.15**	**1.00**	**1**	**0.14**	**114**
PE	+	−	0.74 ±0.30	23.25 ±1.76	23.79 ±1.59	0.29	11	0.00	13
PE*	+	−	0.88 ±0.27	19.41 ±1.58	20.18 ±1.49	0.34	10	0.00	13
PicM	+	−	0.76 ±0.27	18.46 ±1.45	19.05 ±1.31	0.36	9	0.00	194
MMed	+	+	5.79 ±2.06	1.10 e5 ±4026.05	1.11 e5 ±4049.78	0.00	14	1.81	186
MedMad	+	+	0.51 ±0.10	26.73 ±0.74	26.99 ±0.72	0.26	12	6.83	184
kMedMad	+	−	0.59 ±0.07	12.86 ±0.29	13.21 ±0.27	0.52	5	0.63	228
Hybr	+	−	0.62 ±0.07	13.02 ±0.29	13.41 ±0.27	0.51	6	0.00	238
SMLE	+	−	5.14 ±0.06	8.08 ±0.21	34.51 ±0.42	0.20	13	0.00	93
MDE	+	−	0.28 ±0.06	10.18 ±0.24	10.26 ±0.21	0.67	3	0.00	346
MBRE.wc	+	−	0.36 ±0.07	14.17 ±0.33	14.30 ±0.32	0.48	7	0.00	295
OMSE.wc	+	−	0.19 ±0.06	9.44 ±0.21	9.47 ±0.20	0.73	2	0.00	623
MBRE.nc	+	−	0.63 ±0.08	18.19 ±0.43	18.58 ±0.41	0.37	8	0.00	41
OMSE.nc	+	−	0.75 ±0.06	10.68 ±0.28	11.25 ±0.26	0.61	4	0.00	44

contaminated situation:

estimator	s_β	s_ξ	|Bias|	Var	MSE	eff	rank	NA
MLE	−	++	44.16 ±4.82	6.06 e5 ±2.56 e5	6.25 e5 ±2.58 e5	0.00	14	0.14
PE	+	−	3.73 ±0.28	20.74 ±1.49	34.66 ±2.01	0.50	10	0.00
PE*	+	−	4.07 ±0.28	21.08 ±1.62	37.64 ±2.44	0.46	11	0.00
PicM	+	−	2.84 ±0.24	15.27 ±1.10	23.35 ±1.30	0.75	8	0.00
MMed	+	+	3.65 ±1.16	3520.58 ±2275.69	3533.92 ±2283.99	0.00	13	6.25
MedMad	+	−	2.02 ±0.10	26.34 ±0.82	30.42 ±0.68	0.57	9	20.80
kMedMad	+	+	2.11 ±0.08	16.12 ±0.38	20.58 ±0.49	0.85	5	3.71
Hybr	+	+	2.10 ±0.08	16.21 ±0.38	20.64 ±0.49	0.84	6	0.00
SMLE	+	−	0.73 ±0.19	95.38 ±85.59	95.92 ±85.39	0.18	12	0.14
MDE	+	+	2.49 ±0.07	13.34 ±0.32	19.55 ±0.43	0.89	4	0.14
MBRE.wc	−	+	1.75 ±0.08	16.49 ±0.39	19.55 ±0.44	0.89	3	0.00
OMSE.wc	**−**	**+**	**2.43 ±0.07**	**11.52 ±0.28**	**17.41 ±0.39**	**1.00**	**1**	**0.28**
MBRE.nc	+	+	1.54 ±0.09	20.15 ±0.48	22.54 ±0.47	0.77	7	0.00
OMSE.nc	−	+	2.04 ±0.07	14.19 ±0.38	18.33 ±0.43	0.95	2	0.00

Table 9 Comparison of the empirical robustness properties of the estimators at $n = 100$

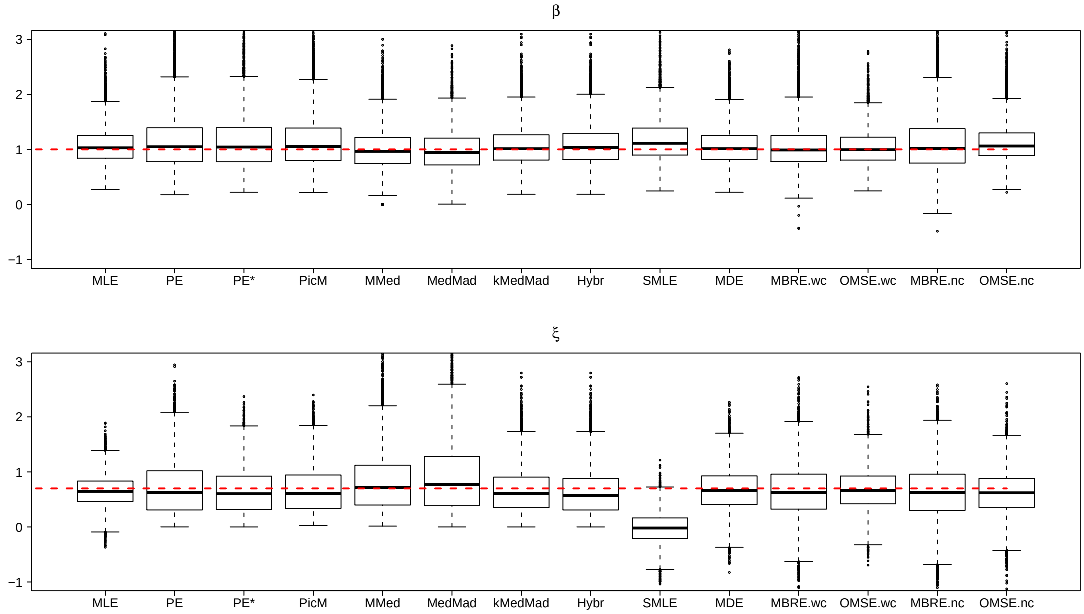
ideal situation:

estimator	s_β	s_ξ	$\lvert\text{Bias}\rvert$	Var	MSE	eff	rank	NA	time
MLE	$+$	$-$	**0.08** ± 0.05	**6.32** ± 0.14	**6.32** ± 0.12	**1.00**	**1**	**0.00**	**396**
PE	$+$	.	0.35 ± 0.30	24.16 ± 1.78	24.29 ± 1.47	0.26	12	0.00	15
PE*	$+$	.	0.30 ± 0.26	17.83 ± 1.42	17.92 ± 1.14	0.35	10	0.00	15
PicM	$+$	$-$	0.34 ± 0.26	17.16 ± 1.31	17.27 ± 1.07	0.37	9	0.00	207
MMed	$+$	.	0.08 ± 0.09	20.14 ± 0.50	20.15 ± 0.43	0.31	11	0.00	260
MedMad	$+$	.	0.11 ± 0.11	30.16 ± 0.76	30.17 ± 0.67	0.21	13	0.00	877
kMedMad	$+$	$-$	0.16 ± 0.07	12.90 ± 0.29	12.93 ± 0.26	0.49	5	0.00	1114
Hybr	$+$	$-$	0.16 ± 0.07	12.90 ± 0.29	12.93 ± 0.26	0.49	5	0.00	1125
SMLE	$+$	$-$	7.66 ± 0.05	7.11 ± 0.16	65.72 ± 0.55	0.10	14	0.00	333
MDE	$+$	.	0.07 ± 0.06	9.82 ± 0.23	9.83 ± 0.20	0.64	3	0.00	564
MBRE.wc	$+$	.	0.08 ± 0.07	13.44 ± 0.29	13.45 ± 0.27	0.47	7	0.00	382
OMSE.wc	.	.	0.03 ± 0.06	9.34 ± 0.21	9.34 ± 0.19	0.68	2	0.00	743
MBRE.nc	$+$	.	0.16 ± 0.08	17.09 ± 0.38	17.12 ± 0.34	0.37	8	0.00	130
OMSE.nc	$+$	$-$	0.23 ± 0.06	9.81 ± 0.23	9.86 ± 0.20	0.64	4	0.00	127

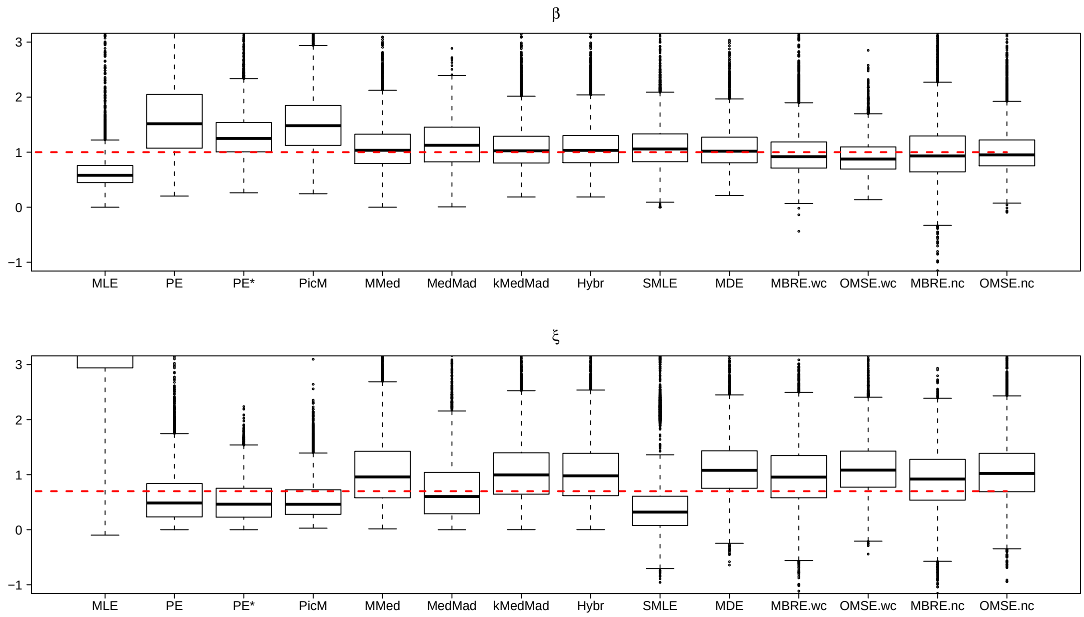
contaminated situation:

estimator	s_β	s_ξ	$\lvert\text{Bias}\rvert$	Var	MSE	eff	rank	NA
MLE	$-$	$++$	32.30 ± 0.12	35.96 ± 0.97	1079.56 ± 7.13	0.01	14	0.00
PE	$+$	$-$	4.44 ± 0.31	25.54 ± 1.91	45.28 ± 2.50	0.33	12	0.00
PE*	$+$	$-$	3.96 ± 0.28	19.96 ± 1.54	35.64 ± 2.01	0.42	11	0.00
PicM	$+$	$-$	2.97 ± 0.26	17.30 ± 1.33	26.13 ± 1.50	0.57	9	0.00
MMed	$+$	$-$	2.81 ± 0.09	19.94 ± 0.48	27.81 ± 0.53	0.53	10	0.00
MedMad	$+$	$-$	3.97 ± 0.11	31.28 ± 0.81	47.07 ± 0.91	0.32	13	0.00
kMedMad	$+$	$+$	2.16 ± 0.07	13.61 ± 0.31	18.28 ± 0.38	0.81	5	0.00
Hybr	$+$	$+$	2.16 ± 0.07	13.61 ± 0.31	18.28 ± 0.38	0.81	5	0.00
SMLE	$+$	$-$	3.00 ± 0.07	12.36 ± 0.61	21.39 ± 0.45	0.69	8	0.00
MDE	$+$	$+$	2.42 ± 0.06	10.68 ± 0.25	16.51 ± 0.33	0.90	3	0.00
MBRE.wc	$-$	$+$	1.82 ± 0.07	14.08 ± 0.31	17.40 ± 0.37	0.85	4	0.00
OMSE.wc	$-$	$+$	2.27 ± 0.06	9.80 ± 0.22	14.96 ± 0.30	0.99	2	0.00
MBRE.nc	$-$	$+$	1.82 ± 0.08	17.66 ± 0.39	20.97 ± 0.42	0.71	7	0.00
OMSE.nc	$-$	$+$	**2.08** ± 0.06	**10.54** ± 0.25	**14.85** ± 0.30	**1.00**	**1**	**0.00**

Table 10: Comparison of the empirical robustness properties of the estimators at $n = 1000$

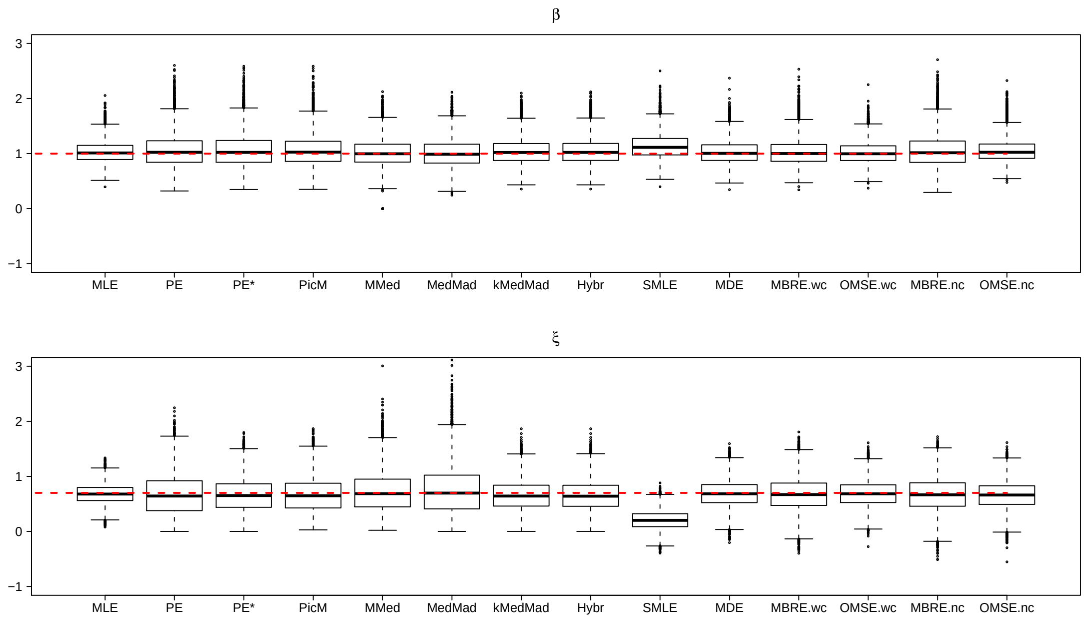


(a) no contamination, 40 sample size

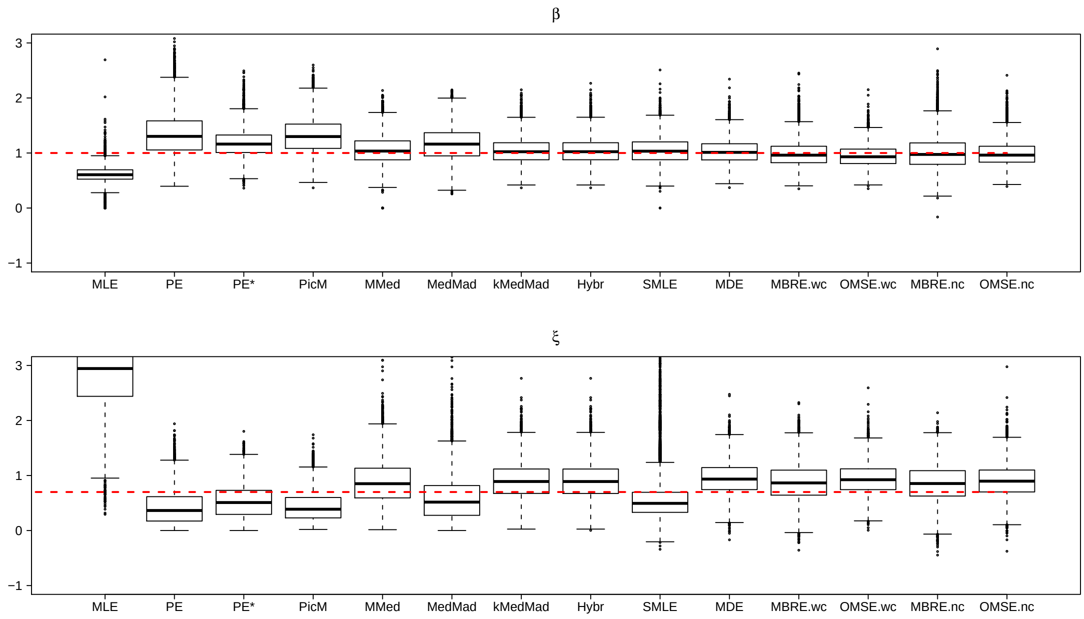


(b) 7.9 % contamination, 40 sample size

Figure 7 Boxplots
for MLE, PE, PE*, PicM, MMed, MedMad, kMedMad, Hybr, SMLE (with $\approx 0.7 \cdot \sqrt{n}$ skipped values), MDE, MBRE.wc, OMSE.wc, MBRE.nc, and OMSE.nc 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

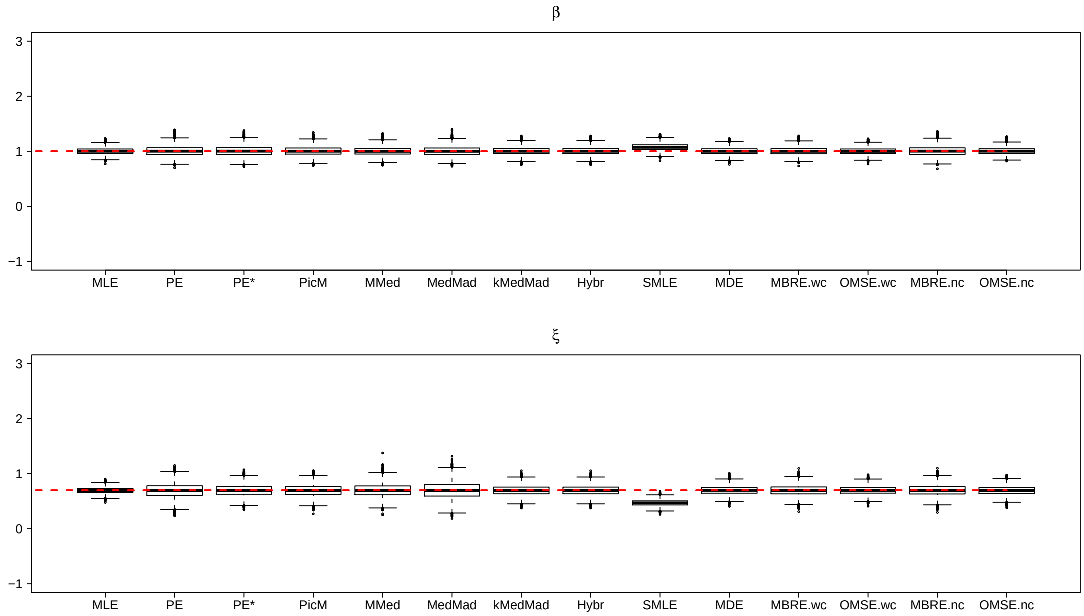


(a) no contamination, 100 sample size

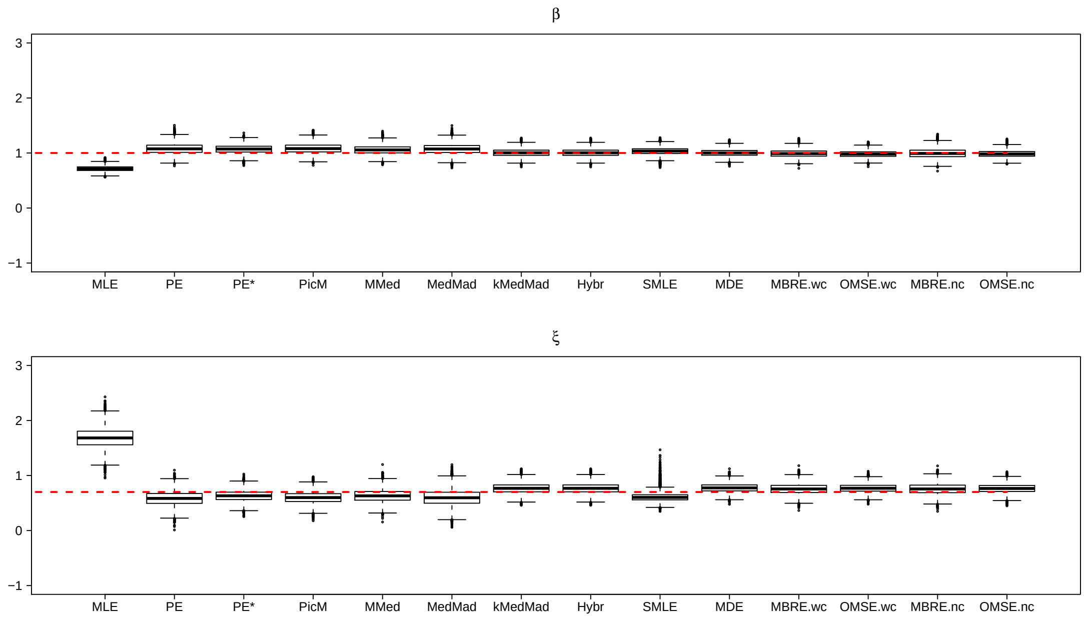


(b) 5 % contamination, 100 sample size

Figure 8: Boxplots—same scale as for $n = 40$
for MLE, PE, PE*, PicM, MMed, MedMad, kMedMad, Hybr, SMLE (with $\approx 0.7 \cdot \sqrt{n}$ skipped values), MDE, MBRE.wc, OMSE.wc, MBRE.nc, and OMSE.nc 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; note the effect of increased sample size as to accuracy.

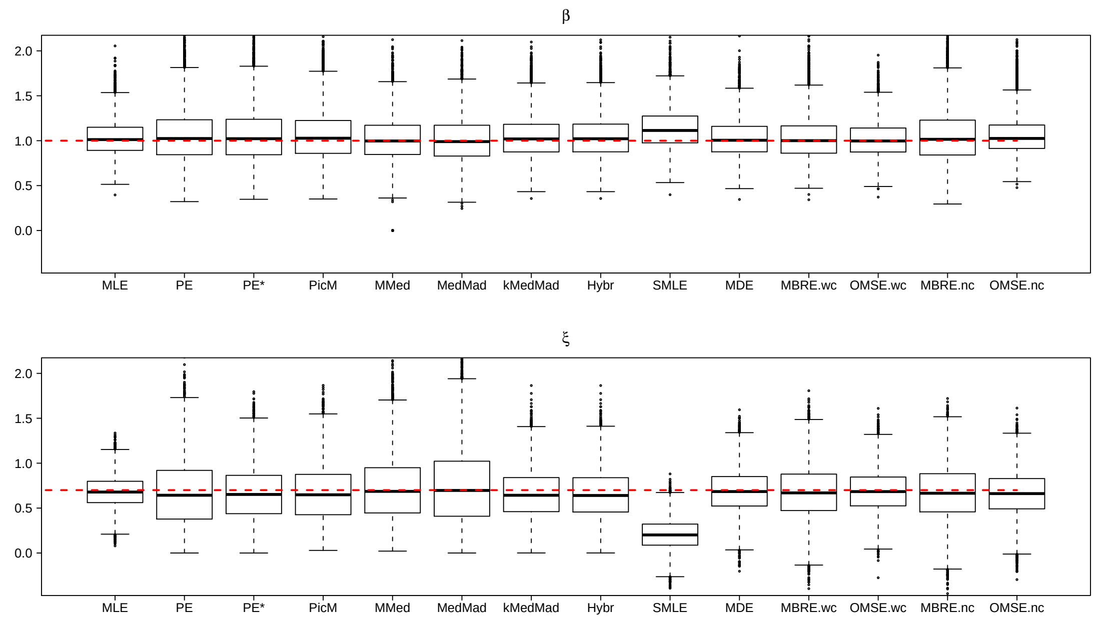


(a) no contamination, 1000 sample size

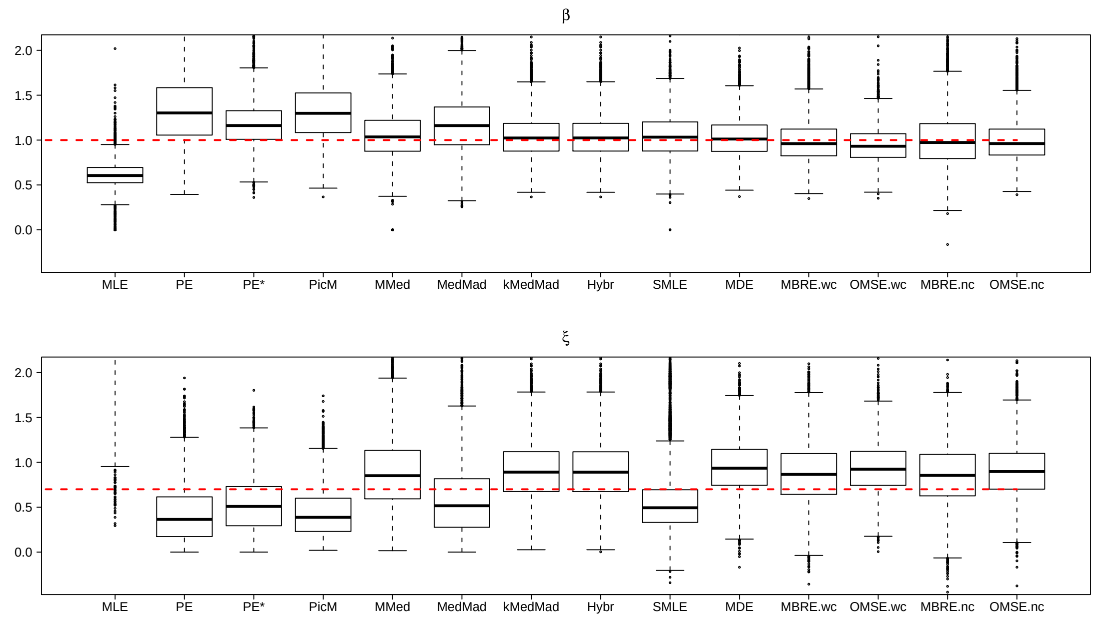


(b) 1.6 % contamination, 1000 sample size

Figure 9: Boxplots—same scale as for $n = 40$
for MLE, PE, PE*, PicM, MMed, MedMad, kMedMad, Hybr, SMLE (with $\approx 0.7 \cdot \sqrt{n}$ skipped values), MDE, MBRE.wc, OMSE.wc, MBRE.nc, and OMSE.nc 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; note the effect of increased sample size as to accuracy.

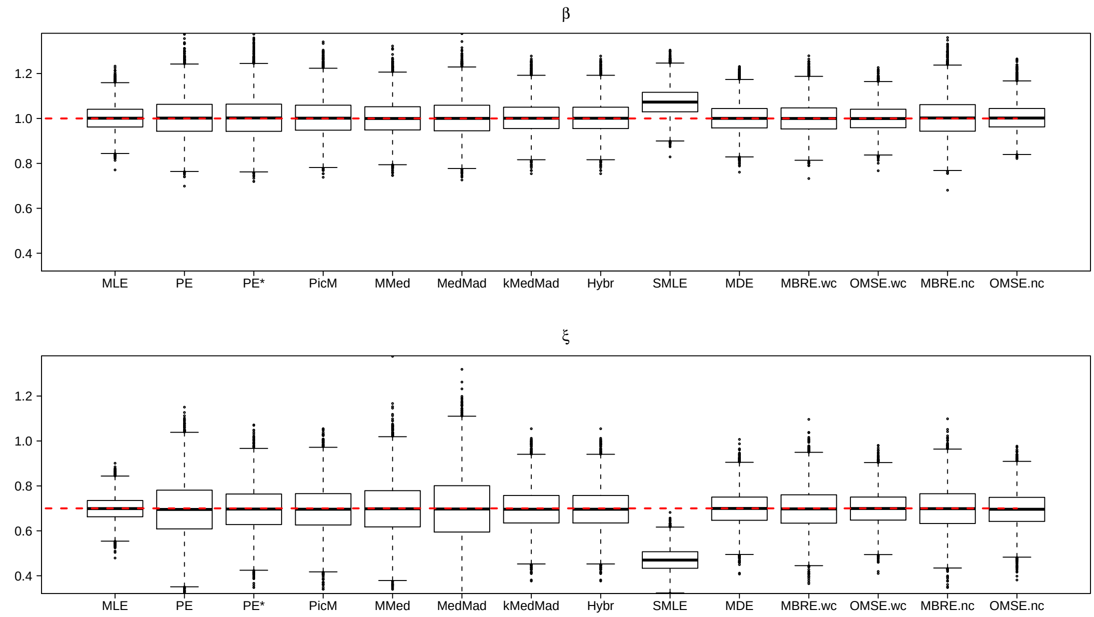


(a) no contamination, 100 sample size

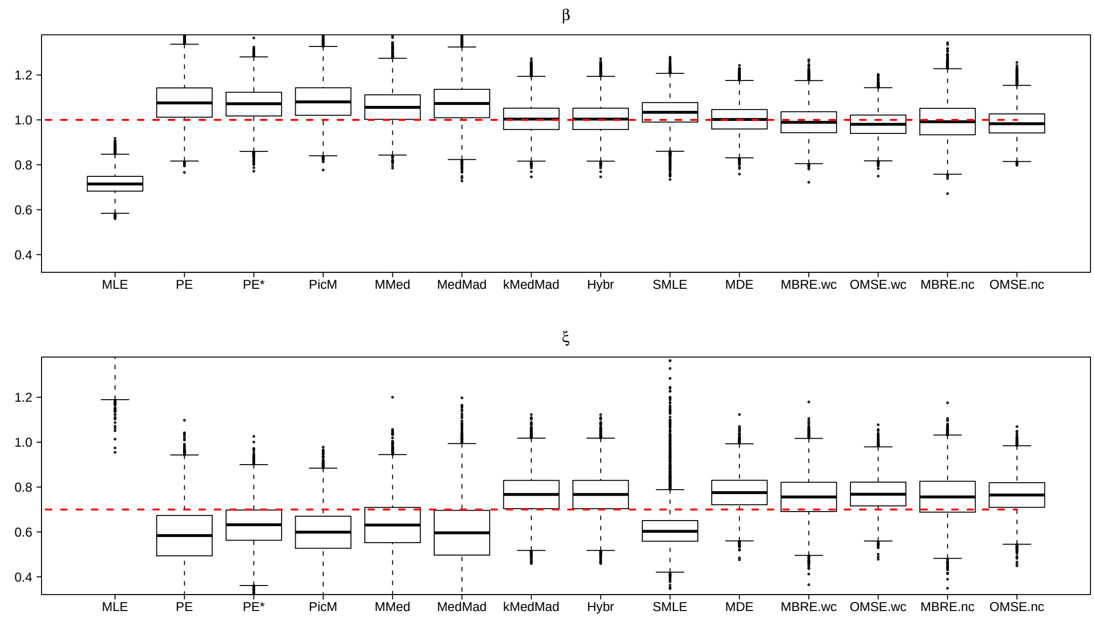


(b) 5 % contamination, 100 sample size

Figure 10: Boxplots (enlarged for better distinction)
for MLE, PE, PE*, PicM, MMed, MedMad, kMedMad, Hybr, SMLE (with $\approx 0.7 \cdot \sqrt{n}$ skipped values), MDE, MBRE.wc, OMSE.wc, MBRE.nc, and OMSE.nc 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



(a) no contamination, 1000 sample size



(b) 1.6 % contamination, 1000 sample size

Figure 11: Boxplots (enlarged to for better distinction)
for MLE, PE, PE*, PicM, MMed, MedMad, kMedMad, Hybr, SMLE (with $\approx 0.7 \cdot \sqrt{n}$ skipped values), MDE, MBRE.wc, OMSE.wc, MBRE.nc, and OMSE.nc 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

Published reports of the Fraunhofer ITWM

The PDF-files of the following reports are available under:
www.itwm.fraunhofer.de/de/ zentral__berichte/berichte

1. D. Hietel, K. Steiner, J. Struckmeier
A Finite - Volume Particle Method for Compressible Flows
(19 pages, 1998)

2. M. Feldmann, S. Seibold
Damage Diagnosis of Rotors: Application of Hilbert Transform and Multi-Hypothesis Testing
Keywords: Hilbert transform, damage diagnosis, Kalman filtering, non-linear dynamics
(23 pages, 1998)

3. Y. Ben-Haim, S. Seibold
Robust Reliability of Diagnostic Multi-Hypothesis Algorithms: Application to Rotating Machinery
Keywords: Robust reliability, convex models, Kalman filtering, multi-hypothesis diagnosis, rotating machinery, crack diagnosis
(24 pages, 1998)

4. F.-Th. Lentes, N. Siedow
Three-dimensional Radiative Heat Transfer in Glass Cooling Processes
(23 pages, 1998)

5. A. Klar, R. Wegener
A hierarchy of models for multilane vehicular traffic
Part I: Modeling
(23 pages, 1998)

Part II: Numerical and stochastic investigations
(17 pages, 1998)

6. A. Klar, N. Siedow
Boundary Layers and Domain Decomposition for Radiative Heat Transfer and Diffusion Equations: Applications to Glass Manufacturing Processes
(24 pages, 1998)

7. I. Choquet
Heterogeneous catalysis modelling and numerical simulation in rarified gas flows Part I: Coverage locally at equilibrium
(24 pages, 1998)

8. J. Ohser, B. Steinbach, C. Lang
Efficient Texture Analysis of Binary Images
(17 pages, 1998)

9. J. Orlik
Homogenization for viscoelasticity of the integral type with aging and shrinkage
(20 pages, 1998)

10. J. Mohring
Helmholtz Resonators with Large Aperture
(21 pages, 1998)

11. H. W. Hamacher, A. Schöbel
On Center Cycles in Grid Graphs
(15 pages, 1998)

12. H. W. Hamacher, K.-H. Küfer
Inverse radiation therapy planning - a multiple objective optimisation approach
(14 pages, 1999)

13. C. Lang, J. Ohser, R. Hilfer
On the Analysis of Spatial Binary Images
(20 pages, 1999)

14. M. Junk
On the Construction of Discrete Equilibrium Distributions for Kinetic Schemes
(24 pages, 1999)

15. M. Junk, S. V. Raghurame Rao
A new discrete velocity method for Navier-Stokes equations
(20 pages, 1999)

16. H. Neunzert
Mathematics as a Key to Key Technologies
(39 pages (4 PDF-Files), 1999)

17. J. Ohser, K. Sandau
Considerations about the Estimation of the Size Distribution in Wicksell's Corpuscle Problem
(18 pages, 1999)

18. E. Carrizosa, H. W. Hamacher, R. Klein, S. Nickel
Solving nonconvex planar location problems by finite dominating sets
Keywords: Continuous Location, Polyhedral Gauges, Finite Dominating Sets, Approximation, Sandwich Algorithm, Greedy Algorithm
(19 pages, 2000)

19. A. Becker
A Review on Image Distortion Measures
Keywords: Distortion measure, human visual system
(26 pages, 2000)

20. H. W. Hamacher, M. Labbé, S. Nickel, T. Sonneborn
Polyhedral Properties of the Uncapacitated Multiple Allocation Hub Location Problem
Keywords: integer programming, hub location, facility location, valid inequalities, facets, branch and cut
(21 pages, 2000)

21. H. W. Hamacher, A. Schöbel
Design of Zone Tariff Systems in Public Transportation
(30 pages, 2001)

22. D. Hietel, M. Junk, R. Keck, D. Teleaga
The Finite-Volume-Particle Method for Conservation Laws
(16 pages, 2001)

23. T. Bender, H. Hennes, J. Kalcsics, M. T. Melo, S. Nickel
Location Software and Interface with GIS and Supply Chain Management
Keywords: facility location, software development, geographical information systems, supply chain management
(48 pages, 2001)

24. H. W. Hamacher, S. A. Tjandra
Mathematical Modelling of Evacuation Problems: A State of Art
(44 pages, 2001)

25. J. Kuhnert, S. Tiwari
Grid free method for solving the Poisson equation
Keywords: Poisson equation, Least squares method, Grid free method
(19 pages, 2001)

26. T. Götz, H. Rave, D. Reinel-Bitzer, K. Steiner, H. Tiemeier
Simulation of the fiber spinning process
Keywords: Melt spinning, fiber model, Lattice Boltzmann, CFD
(19 pages, 2001)

27. A. Zemitis
On interaction of a liquid film with an obstacle
Keywords: impinging jets, liquid film, models, numerical solution, shape
(22 pages, 2001)

28. I. Ginzburg, K. Steiner
Free surface lattice-Boltzmann method to model the filling of expanding cavities by Bingham Fluids
Keywords: Generalized LBE, free-surface phenomena, interface boundary conditions, filling processes, Bingham viscoplastic model, regularized models
(22 pages, 2001)

29. H. Neunzert
»Denn nichts ist für den Menschen als Menschen etwas wert, was er nicht mit Leidenschaft tun kann«
Vortrag anlässlich der Verleihung des Akademiepreises des Landes Rheinland-Pfalz am 21.11.2001
Keywords: Lehre, Forschung, angewandte Mathematik, Mehrskalenanalyse, Strömungsmechanik
(18 pages, 2001)

30. J. Kuhnert, S. Tiwari
Finite pointset method based on the projection method for simulations of the incompressible Navier-Stokes equations
Keywords: Incompressible Navier-Stokes equations, Meshfree method, Projection method, Particle scheme, Least squares approximation
AMS subject classification: 76D05, 76M28
(25 pages, 2001)

31. R. Korn, M. Krekel
Optimal Portfolios with Fixed Consumption or Income Streams
Keywords: Portfolio optimisation, stochastic control, HJB equation, discretisation of control problems
(23 pages, 2002)

32. M. Krekel
Optimal portfolios with a loan dependent credit spread
Keywords: Portfolio optimisation, stochastic control, HJB equation, credit spread, log utility, power utility, non-linear wealth dynamics
(25 pages, 2002)

33. J. Ohser, W. Nagel, K. Schladitz
The Euler number of discretized sets – on the choice of adjacency in homogeneous lattices
Keywords: image analysis, Euler number, neighborhod relationships, cuboidal lattice
(32 pages, 2002)

34. I. Ginzburg, K. Steiner
Lattice Boltzmann Model for Free-Surface flow and Its Application to Filling Process in Casting
Keywords: Lattice Boltzmann models; free-surface phenomena; interface boundary conditions; filling processes; injection molding; volume of fluid method; interface boundary conditions; advection-schemes; upwind-schemes
(54 pages, 2002)

35. M. Günther, A. Klar, T. Materne, R. Wegener
Multivalued fundamental diagrams and stop and go waves for continuum traffic equations
Keywords: traffic flow, macroscopic equations, kinetic derivation, multivalued fundamental diagram, stop and go waves, phase transitions
(25 pages, 2002)

36. S. Feldmann, P. Lang, D. Prätzel-Wolters
Parameter influence on the zeros of network determinants
Keywords: Networks, Equicofactor matrix polynomials, Realization theory, Matrix perturbation theory
(30 pages, 2002)

37. K. Koch, J. Ohser, K. Schladitz
Spectral theory for random closed sets and estimating the covariance via frequency space
Keywords: Random set, Bartlett spectrum, fast Fourier transform, power spectrum
(28 pages, 2002)

38. D. d'Humières, I. Ginzburg
Multi-reflection boundary conditions for lattice Boltzmann models
Keywords: lattice Boltzmann equation, boudary condistions, bounce-back rule, Navier-Stokes equation
(72 pages, 2002)

39. R. Korn
Elementare Finanzmathematik
Keywords: Finanzmathematik, Aktien, Optionen, Portfolio-Optimierung, Börse, Lehrerweiterbildung, Mathematikunterricht
(98 pages, 2002)

40. J. Kallrath, M. C. Müller, S. Nickel
Batch Presorting Problems: Models and Complexity Results
Keywords: Complexity theory, Integer programming, Assigment, Logistics
(19 pages, 2002)

41. J. Linn
On the frame-invariant description of the phase space of the Folgar-Tucker equation
Key words: fiber orientation, Folgar-Tucker equation, injection molding
(5 pages, 2003)

42. T. Hanne, S. Nickel
A Multi-Objective Evolutionary Algorithm for Scheduling and Inspection Planning in Software Development Projects
Key words: multiple objective programming, project management and scheduling, software development, evolutionary algorithms, efficient set
(29 pages, 2003)

43. T. Bortfeld , K.-H. Küfer, M. Monz, A. Scherrer, C. Thieke, H. Trinkaus
Intensity-Modulated Radiotherapy - A Large Scale Multi-Criteria Programming Problem
Keywords: multiple criteria optimization, representative systems of Pareto solutions, adaptive triangulation, clustering and disaggregation techniques, visualization of Pareto solutions, medical physics, external beam radiotherapy planning, intensity modulated radiotherapy
(31 pages, 2003)

44. T. Halfmann, T. Wichmann
Overview of Symbolic Methods in Industrial Analog Circuit Design
Keywords: CAD, automated analog circuit design, symbolic analysis, computer algebra, behavioral modeling, system simulation, circuit sizing, macro modeling, differential-algebraic equations, index
(17 pages, 2003)

45. S. E. Mikhailov, J. Orlik
Asymptotic Homogenisation in Strength and Fatigue Durability Analysis of Composites
Keywords: multiscale structures, asymptotic homogenization, strength, fatigue, singularity, non-local conditions
(14 pages, 2003)

46. P. Domínguez-Marín, P. Hansen, N. Mladenovi´c , S. Nickel
Heuristic Procedures for Solving the Discrete Ordered Median Problem
Keywords: genetic algorithms, variable neighborhood search, discrete facility location
(31 pages, 2003)

47. N. Boland, P. Domínguez-Marín, S. Nickel, J. Puerto
Exact Procedures for Solving the Discrete Ordered Median Problem
Keywords: discrete location, Integer programming
(41 pages, 2003)

48. S. Feldmann, P. Lang
Padé-like reduction of stable discrete linear systems preserving their stability
Keywords: Discrete linear systems, model reduction, stability, Hankel matrix, Stein equation
(16 pages, 2003)

49. J. Kallrath, S. Nickel
A Polynomial Case of the Batch Presorting Problem
Keywords: batch presorting problem, online optimization, competetive analysis, polynomial algorithms, logistics
(17 pages, 2003)

50. T. Hanne, H. L. Trinkaus
knowCube for MCDM – Visual and Interactive Support for Multicriteria Decision Making
Key words: Multicriteria decision making, knowledge management, decision support systems, visual interfaces, interactive navigation, real-life applications.
(26 pages, 2003)

51. O. Iliev, V. Laptev
On Numerical Simulation of Flow Through Oil Filters
Keywords: oil filters, coupled flow in plain and porous media, Navier-Stokes, Brinkman, numerical simulation
(8 pages, 2003)

52. W. Dörfler, O. Iliev, D. Stoyanov, D. Vassileva
On a Multigrid Adaptive Refinement Solver for Saturated Non-Newtonian Flow in Porous Media
Keywords: Nonlinear multigrid, adaptive refinement, non-Newtonian flow in porous media
(17 pages, 2003)

53. S. Kruse
On the Pricing of Forward Starting Options under Stochastic Volatility
Keywords: Option pricing, forward starting options, Heston model, stochastic volatility, cliquet options
(11 pages, 2003)

54. O. Iliev, D. Stoyanov
Multigrid – adaptive local refinement solver for incompressible flows
Keywords: Navier-Stokes equations, incompressible flow, projection-type splitting, SIMPLE, multigrid methods, adaptive local refinement, lid-driven flow in a cavity
(37 pages, 2003)

55. V. Starikovicius
The multiphase flow and heat transfer in porous media
Keywords: Two-phase flow in porous media, various formulations, global pressure, multiphase mixture model, numerical simulation
(30 pages, 2003)

56. P. Lang, A. Sarishvili, A. Wirsen
Blocked neural networks for knowledge extraction in the software development process
Keywords: Blocked Neural Networks, Nonlinear Regression, Knowledge Extraction, Code Inspection
(21 pages, 2003)

57. H. Knaf, P. Lang, S. Zeiser
Diagnosis aiding in Regulation Thermography using Fuzzy Logic
Keywords: fuzzy logic,knowledge representation, expert system
(22 pages, 2003)

58. M. T. Melo, S. Nickel, F. Saldanha da Gama
Largescale models for dynamic multicommodity capacitated facility location
Keywords: supply chain management, strategic planning, dynamic location, modeling
(40 pages, 2003)

59. J. Orlik
Homogenization for contact problems with periodically rough surfaces
Keywords: asymptotic homogenization, contact problems
(28 pages, 2004)

60. A. Scherrer, K.-H. Küfer, M. Monz, F. Alonso, T. Bortfeld
IMRT planning on adaptive volume structures – a significant advance of computational complexity
Keywords: Intensity-modulated radiation therapy (IMRT), inverse treatment planning, adaptive volume structures, hierarchical clustering, local refinement, adaptive clustering, convex programming, mesh generation, multi-grid methods
(24 pages, 2004)

61. D. Kehrwald
Parallel lattice Boltzmann simulation of complex flows
Keywords: Lattice Boltzmann methods, parallel computing, microstructure simulation, virtual material design, pseudo-plastic fluids, liquid composite moulding
(12 pages, 2004)

62. O. Iliev, J. Linn, M. Moog, D. Niedziela, V. Starikovicius
On the Performance of Certain Iterative Solvers for Coupled Systems Arising in Discretization of Non-Newtonian Flow Equations

Keywords: Performance of iterative solvers, Preconditioners, Non-Newtonian flow
(17 pages, 2004)

63. R. Ciegis, O. Iliev, S. Rief, K. Steiner
On Modelling and Simulation of Different Regimes for Liquid Polymer Moulding
Keywords: Liquid Polymer Moulding, Modelling, Simulation, Infiltration, Front Propagation, non-Newtonian flow in porous media
(43 pages, 2004)

64. T. Hanne, H. Neu
Simulating Human Resources in Software Development Processes
Keywords: Human resource modeling, software process, productivity, human factors, learning curve
(14 pages, 2004)

65. O. Iliev, A. Mikelic, P. Popov
Fluid structure interaction problems in deformable porous media: Toward permeability of deformable porous media
Keywords: fluid-structure interaction, deformable porous media, upscaling, linear elasticity, stokes, finite elements
(28 pages, 2004)

66. F. Gaspar, O. Iliev, F. Lisbona, A. Naumovich, P. Vabishchevich
On numerical solution of 1-D poroelasticity equations in a multilayered domain
Keywords: poroelasticity, multilayered material, finite volume discretization, MAC type grid
(41 pages, 2004)

67. J. Ohser, K. Schladitz, K. Koch, M. Nöthe
Diffraction by image processing and its application in materials science
Keywords: porous microstructure, image analysis, random set, fast Fourier transform, power spectrum, Bartlett spectrum
(13 pages, 2004)

68. H. Neunzert
Mathematics as a Technology: Challenges for the next 10 Years
Keywords: applied mathematics, technology, modelling, simulation, visualization, optimization, glass processing, spinning processes, fiber-fluid interaction, trubulence effects, topological optimization, multicriteria optimization, Uncertainty and Risk, financial mathematics, Malliavin calculus, Monte-Carlo methods, virtual material design, filtration, bio-informatics, system biology
(29 pages, 2004)

69. R. Ewing, O. Iliev, R. Lazarov, A. Naumovich
On convergence of certain finite difference discretizations for 1D poroelasticity interface problems
Keywords: poroelasticity, multilayered material, finite volume discretizations, MAC type grid, error estimates
(26 pages,2004)

70. W. Dörfler, O. Iliev, D. Stoyanov, D. Vassileva
On Efficient Simulation of Non-Newtonian Flow in Saturated Porous Media with a Multigrid Adaptive Refinement Solver
Keywords: Nonlinear multigrid, adaptive renement, non-Newtonian in porous media
(25 pages, 2004)

71. J. Kalcsics, S. Nickel, M. Schröder
Towards a Unified Territory Design Approach – Applications, Algorithms and GIS Integration
Keywords: territory desgin, political districting, sales territory alignment, optimization algorithms, Geographical Information Systems
(40 pages, 2005)

72. K. Schladitz, S. Peters, D. Reinel-Bitzer, A. Wiegmann, J. Ohser
Design of acoustic trim based on geometric modeling and flow simulation for non-woven
Keywords: random system of fibers, Poisson line process, flow resistivity, acoustic absorption, Lattice-Boltzmann method, non-woven
(21 pages, 2005)

73. V. Rutka, A. Wiegmann
Explicit Jump Immersed Interface Method for virtual material design of the effective elastic moduli of composite materials
Keywords: virtual material design, explicit jump immersed interface method, effective elastic moduli, composite materials
(22 pages, 2005)

74. T. Hanne
Eine Übersicht zum Scheduling von Baustellen
Keywords: Projektplanung, Scheduling, Bauplanung, Bauindustrie
(32 pages, 2005)

75. J. Linn
The Folgar-Tucker Model as a Differetial Algebraic System for Fiber Orientation Calculation
Keywords: fiber orientation, Folgar–Tucker model, invariants, algebraic constraints, phase space, trace stability
(15 pages, 2005)

76. M. Speckert, K. Dreßler, H. Mauch, A. Lion, G. J. Wierda
Simulation eines neuartigen Prüfsystems für Achserprobungen durch MKS-Modellierung einschließlich Regelung
Keywords: virtual test rig, suspension testing, multibody simulation, modeling hexapod test rig, optimization of test rig configuration
(20 pages, 2005)

77. K.-H. Küfer, M. Monz, A. Scherrer, P. Süss, F. Alonso, A. S. A. Sultan, Th. Bortfeld, D. Craft, Chr. Thieke
Multicriteria optimization in intensity modulated radiotherapy planning
Keywords: multicriteria optimization, extreme solutions, real-time decision making, adaptive approximation schemes, clustering methods, IMRT planning, reverse engineering
(51 pages, 2005)

78. S. Amstutz, H. Andrä
A new algorithm for topology optimization using a level-set method
Keywords: shape optimization, topology optimization, topological sensitivity, level-set
(22 pages, 2005)

79. N. Ettrich
Generation of surface elevation models for urban drainage simulation
Keywords: Flooding, simulation, urban elevation models, laser scanning
(22 pages, 2005)

80. H. Andrä, J. Linn, I. Matei, I. Shklyar, K. Steiner, E. Teichmann
OPTCAST – Entwicklung adäquater Strukturoptimierungsverfahren für Gießereien Technischer Bericht (KURZFASSUNG)
Keywords: Topologieoptimierung, Level-Set-Methode, Gießprozesssimulation, Gießtechnische Restriktionen, CAE-Kette zur Strukturoptimierung
(77 pages, 2005)

81. N. Marheineke, R. Wegener
Fiber Dynamics in Turbulent Flows
Part I: General Modeling Framework
Keywords: fiber-fluid interaction; Cosserat rod; turbulence modeling; Kolmogorov's energy spectrum; double-velocity correlations; differentiable Gaussian fields
(20 pages, 2005)

Part II: Specific Taylor Drag
Keywords: flexible fibers; k-ε turbulence model; fiber-turbulence interaction scales; air drag; random Gaussian aerodynamic force; white noise; stochastic differential equations; ARMA process
(18 pages, 2005)

82. C. H. Lampert, O. Wirjadi
An Optimal Non-Orthogonal Separation of the Anisotropic Gaussian Convolution Filter
Keywords: Anisotropic Gaussian filter, linear filtering, orientation space, nD image processing, separable filters
(25 pages, 2005)

83. H. Andrä, D. Stoyanov
Error indicators in the parallel finite element solver for linear elasticity DDFEM
Keywords: linear elasticity, finite element method, hierarchical shape functions, domain decom-position, parallel implementation, a posteriori error estimates
(21 pages, 2006)

84. M. Schröder, I. Solchenbach
Optimization of Transfer Quality in Regional Public Transit
Keywords: public transit, transfer quality, quadratic assignment problem
(16 pages, 2006)

85. A. Naumovich, F. J. Gaspar
On a multigrid solver for the three-dimensional Biot poroelasticity system in multilayered domains
Keywords: poroelasticity, interface problem, multigrid, operator-dependent prolongation
(11 pages, 2006)

86. S. Panda, R. Wegener, N. Marheineke
Slender Body Theory for the Dynamics of Curved Viscous Fibers
Keywords: curved viscous fibers; fluid dynamics; Navier-Stokes equations; free boundary value problem; asymptotic expansions; slender body theory
(14 pages, 2006)

87. E. Ivanov, H. Andrä, A. Kudryavtsev
Domain Decomposition Approach for Automatic Parallel Generation of Tetrahedral Grids
Key words: Grid Generation, Unstructured Grid, Delaunay Triangulation, Parallel Programming, Domain Decomposition, Load Balancing
(18 pages, 2006)

88. S. Tiwari, S. Antonov, D. Hietel, J. Kuhnert, R. Wegener
A Meshfree Method for Simulations of Interactions between Fluids and Flexible Structures
Key words: Meshfree Method, FPM, Fluid Structure Interaction, Sheet of Paper, Dynamical Coupling
(16 pages, 2006)

89. R. Ciegis , O. Iliev, V. Starikovicius, K. Steiner
Numerical Algorithms for Solving Problems of Multiphase Flows in Porous Media
Keywords: nonlinear algorithms, finite-volume method, software tools, porous media, flows
(16 pages, 2006)

Keywords: numerical upscaling, heterogeneous porous media, single phase flow, Darcy's law, multiscale problem, effective permeability, multipoint flux approximation, anisotropy
(17 pages, 2007)

119. O. Iliev, I. Rybak
On approximation property of multipoint flux approximation method
Keywords: Multipoint flux approximation, finite volume method, elliptic equation, discontinuous tensor coefficients, anisotropy
(15 pages, 2007)

120. O. Iliev, I. Rybak, J. Willems
On upscaling heat conductivity for a class of industrial problems
Keywords: Multiscale problems, effective heat conductivity, numerical upscaling, domain decomposition
(21 pages, 2007)

121. R. Ewing, O. Iliev, R. Lazarov, I. Rybak
On two-level preconditioners for flow in porous media
Keywords: Multiscale problem, Darcy's law, single phase flow, anisotropic heterogeneous porous media, numerical upscaling, multigrid, domain decomposition, efficient preconditioner
(18 pages, 2007)

122. M. Brickenstein, A. Dreyer
POLYBORI: A Gröbner basis framework for Boolean polynomials
Keywords: Gröbner basis, formal verification, Boolean polynomials, algebraic cryptoanalysis, satisfiability
(23 pages, 2007)

123. O. Wirjadi
Survey of 3d image segmentation methods
Keywords: image processing, 3d, image segmentation, binarization
(20 pages, 2007)

124. S. Zeytun, A. Gupta
A Comparative Study of the Vasicek and the CIR Model of the Short Rate
Keywords: interest rates, Vasicek model, CIR-model, calibration, parameter estimation
(17 pages, 2007)

125. G. Hanselmann, A. Sarishvili
Heterogeneous redundancy in software quality prediction using a hybrid Bayesian approach
Keywords: reliability prediction, fault prediction, non-homogeneous poisson process, Bayesian model averaging
(17 pages, 2007)

126. V. Maag, M. Berger, A. Winterfeld, K.-H. Küfer
A novel non-linear approach to minimal area rectangular packing
Keywords: rectangular packing, non-overlapping constraints, non-linear optimization, regularization, relaxation
(18 pages, 2007)

127. M. Monz, K.-H. Küfer, T. Bortfeld, C. Thieke
Pareto navigation – systematic multi-criteria-based IMRT treatment plan determination
Keywords: convex, interactive multi-objective optimization, intensity modulated radiotherapy planning
(15 pages, 2007)

128. M. Krause, A. Scherrer
On the role of modeling parameters in IMRT plan optimization
Keywords: intensity-modulated radiotherapy (IMRT), inverse IMRT planning, convex optimization, sensitivity analysis, elasticity, modeling parameters, equivalent uniform dose (EUD)
(18 pages, 2007)

129. A. Wiegmann
Computation of the permeability of porous materials from their microstructure by FFF-Stokes
Keywords: permeability, numerical homogenization, fast Stokes solver
(24 pages, 2007)

130. T. Melo, S. Nickel, F. Saldanha da Gama
Facility Location and Supply Chain Management – A comprehensive review
Keywords: facility location, supply chain management, network design
(54 pages, 2007)

131. T. Hanne, T. Melo, S. Nickel
Bringing robustness to patient flow management through optimized patient transports in hospitals
Keywords: Dial-a-Ride problem, online problem, case study, tabu search, hospital logistics
(23 pages, 2007)

132. R. Ewing, O. Iliev, R. Lazarov, I. Rybak, J. Willems
An efficient approach for upscaling properties of composite materials with high contrast of coefficients
Keywords: effective heat conductivity, permeability of fractured porous media, numerical upscaling, fibrous insulation materials, metal foams
(16 pages, 2008)

133. S. Gelareh, S. Nickel
New approaches to hub location problems in public transport planning
Keywords: integer programming, hub location, transportation, decomposition, heuristic
(25 pages, 2008)

134. G. Thömmes, J. Becker, M. Junk, A. K. Vaikuntam, D. Kehrwald, A. Klar, K. Steiner, A. Wiegmann
A Lattice Boltzmann Method for immiscible multiphase flow simulations using the Level Set Method
Keywords: Lattice Boltzmann method, Level Set method, free surface, multiphase flow
(28 pages, 2008)

135. J. Orlik
Homogenization in elasto-plasticity
Keywords: multiscale structures, asymptotic homogenization, nonlinear energy
(40 pages, 2008)

136. J. Almquist, H. Schmidt, P. Lang, J. Deitmer, M. Jirstrand, D. Prätzel-Wolters, H. Becker
Determination of interaction between MCT1 and CAII via a mathematical and physiological approach
Keywords: mathematical modeling; model reduction; electrophysiology; pH-sensitive microelectrodes; proton antenna
(20 pages, 2008)

137. E. Savenkov, H. Andrä, O. Iliev
An analysis of one regularization approach for solution of pure Neumann problem
Keywords: pure Neumann problem, elasticity, regularization, finite element method, condition number
(27 pages, 2008)

138. O. Berman, J. Kalcsics, D. Krass, S. Nickel
The ordered gradual covering location problem on a network
Keywords: gradual covering, ordered median function, network location
(32 pages, 2008)

139. S. Gelareh, S. Nickel
Multi-period public transport design: A novel model and solution approaches
Keywords: Integer programming, hub location, public transport, multi-period planning, heuristics
(31 pages, 2008)

140. T. Melo, S. Nickel, F. Saldanha-da-Gama
Network design decisions in supply chain planning
Keywords: supply chain design, integer programming models, location models, heuristics
(20 pages, 2008)

141. C. Lautensack, A. Särkkä, J. Freitag, K. Schladitz
Anisotropy analysis of pressed point processes
Keywords: estimation of compression, isotropy test, nearest neighbour distance, orientation analysis, polar ice, Ripley's K function
(35 pages, 2008)

142. O. Iliev, R. Lazarov, J. Willems
A Graph-Laplacian approach for calculating the effective thermal conductivity of complicated fiber geometries
Keywords: graph laplacian, effective heat conductivity, numerical upscaling, fibrous materials
(14 pages, 2008)

143. J. Linn, T. Stephan, J. Carlsson, R. Bohlin
Fast simulation of quasistatic rod deformations for VR applications
Keywords: quasistatic deformations, geometrically exact rod models, variational formulation, energy minimization, finite differences, nonlinear conjugate gradients
(7 pages, 2008)

144. J. Linn, T. Stephan
Simulation of quasistatic deformations using discrete rod models
Keywords: quasistatic deformations, geometrically exact rod models, variational formulation, energy minimization, finite differences, nonlinear conjugate gradients
(9 pages, 2008)

145. J. Marburger, N. Marheineke, R. Pinnau
Adjoint based optimal control using meshless discretizations
Keywords: Mesh-less methods, particle methods, Eulerian-Lagrangian formulation, optimization strategies, adjoint method, hyperbolic equations
(14 pages, 2008

146. S. Desmettre, J. Gould, A. Szimayer
Own-company stockholding and work effort preferences of an unconstrained executive
Keywords: optimal portfolio choice, executive compensation
(33 pages, 2008)

147. M. Berger, M. Schröder, K.-H. Küfer
A constraint programming approach for the two-dimensional rectangular packing problem with orthogonal orientations
Keywords: rectangular packing, orthogonal orientations non-overlapping constraints, constraint propagation
(13 pages, 2008)

148. K. Schladitz, C. Redenbach, T. Sych, M. Godehardt
Microstructural characterisation of open foams using 3d images
Keywords: virtual material design, image analysis, open foams
(30 pages, 2008)

149. E. Fernández, J. Kalcsics, S. Nickel, R. Ríos-Mercado
A novel territory design model arising in the implementation of the WEEE-Directive
Keywords: heuristics, optimization, logistics, recycling
(28 pages, 2008)

150. H. Lang, J. Linn
Lagrangian field theory in space-time for geometrically exact Cosserat rods
Keywords: Cosserat rods, geometrically exact rods, small strain, large deformation, deformable bodies, Lagrangian field theory, variational calculus
(19 pages, 2009)

151. K. Dreßler, M. Speckert, R. Müller, Ch. Weber
Customer loads correlation in truck engineering
Keywords: Customer distribution, safety critical components, quantile estimation, Monte-Carlo methods
(11 pages, 2009)

152. H. Lang, K. Dreßler
An improved multiaxial stress-strain correction model for elastic FE postprocessing
Keywords: Jiang's model of elastoplasticity, stress-strain correction, parameter identification, automatic differentiation, least-squares optimization, Coleman-Li algorithm
(6 pages, 2009)

153. J. Kalcsics, S. Nickel, M. Schröder
A generic geometric approach to territory design and districting
Keywords: Territory design, districting, combinatorial optimization, heuristics, computational geometry
(32 pages, 2009)

154. Th. Fütterer, A. Klar, R. Wegener
An energy conserving numerical scheme for the dynamics of hyperelastic rods
Keywords: Cosserat rod, hyperealstic, energy conservation, finite differences
(16 pages, 2009)

155. A. Wiegmann, L. Cheng, E. Glatt, O. Iliev, S. Rief
Design of pleated filters by computer simulations
Keywords: Solid-gas separation, solid-liquid separation, pleated filter, design, simulation
(21 pages, 2009)

156. A. Klar, N. Marheineke, R. Wegener
Hierarchy of mathematical models for production processes of technical textiles
Keywords: Fiber-fluid interaction, slender-body theory, turbulence modeling, model reduction, stochastic differential equations, Fokker-Planck equation, asymptotic expansions, parameter identification
(21 pages, 2009)

157. E. Glatt, S. Rief, A. Wiegmann, M. Knefel, E. Wegenke
Structure and pressure drop of real and virtual metal wire meshes
Keywords: metal wire mesh, structure simulation, model calibration, CFD simulation, pressure loss
(7 pages, 2009)

158. S. Kruse, M. Müller
Pricing American call options under the assumption of stochastic dividends – An application of the Korn-Rogers model
Keywords: option pricing, American options, dividends, dividend discount model, Black-Scholes model
(22 pages, 2009)

159. H. Lang, J. Linn, M. Arnold
Multibody dynamics simulation of geometrically exact Cosserat rods
Keywords: flexible multibody dynamics, large deformations, finite rotations, constrained mechanical systems, structural dynamics
(20 pages, 2009)

160. P. Jung, S. Leyendecker, J. Linn, M. Ortiz
Discrete Lagrangian mechanics and geometrically exact Cosserat rods
Keywords: special Cosserat rods, Lagrangian mechanics, Noether's theorem, discrete mechanics, frame-indifference, holonomic constraints
(14 pages, 2009)

161. M. Burger, K. Dreßler, A. Marquardt, M. Speckert
Calculating invariant loads for system simulation in vehicle engineering
Keywords: iterative learning control, optimal control theory, differential algebraic equations(DAEs)
(18 pages, 2009)

162. M. Speckert, N. Ruf, K. Dreßler
Undesired drift of multibody models excited by measured accelerations or forces
Keywords: multibody simulation, full vehicle model, force-based simulation, drift due to noise
(19 pages, 2009)

163. A. Streit, K. Dreßler, M. Speckert, J. Lichter, T. Zenner, P. Bach
Anwendung statistischer Methoden zur Erstellung von Nutzungsprofilen für die Auslegung von Mobilbaggern
Keywords: Nutzungsvielfalt, Kundenbeanspruchung, Bemessungsgrundlagen
(13 pages, 2009)

164. I. Correia, S. Nickel, F. Saldanha-da-Gama
Anwendung statistischer Methoden zur Erstellung von Nutzungsprofilen für die Auslegung von Mobilbaggern
Keywords: Capacitated Hub Location, MIP formulations
(10 pages, 2009)

165. F. Yaneva, T. Grebe, A. Scherrer
An alternative view on global radiotherapy optimization problems
Keywords: radiotherapy planning, path-connected sublevelsets, modified gradient projection method, improving and feasible directions
(14 pages, 2009)

166. J. I. Serna, M. Monz, K.-H. Küfer, C. Thieke
Trade-off bounds and their effect in multi-criteria IMRT planning
Keywords: trade-off bounds, multi-criteria optimization, IMRT, Pareto surface
(15 pages, 2009)

167. W. Arne, N. Marheineke, A. Meister, R. Wegener
Numerical analysis of Cosserat rod and string models for viscous jets in rotational spinning processes
Keywords: Rotational spinning process, curved viscous fibers, asymptotic Cosserat models, boundary value problem, existence of numerical solutions
(18 pages, 2009)

168. T. Melo, S. Nickel, F. Saldanha-da-Gama
An LP-rounding heuristic to solve a multi-period facility relocation problem
Keywords: supply chain design, heuristic, linear programming, rounding
(37 pages, 2009)

169. I. Correia, S. Nickel, F. Saldanha-da-Gama
Single-allocation hub location problems with capacity choices
Keywords: hub location, capacity decisions, MILP formulations
(27 pages, 2009)

170. S. Acar, K. Natcheva-Acar
A guide on the implementation of the Heath-Jarrow-Morton Two-Factor Gaussian Short Rate Model (HJM-G2++)
Keywords: short rate model, two factor Gaussian, G2++, option pricing, calibration
(30 pages, 2009)

171. A. Szimayer, G. Dimitroff, S. Lorenz
A parsimonious multi-asset Heston model: calibration and derivative pricing
Keywords: Heston model, multi-asset, option pricing, calibration, correlation
(28 pages, 2009)

172. N. Marheineke, R. Wegener
Modeling and validation of a stochastic drag for fibers in turbulent flows
Keywords: fiber-fluid interactions, long slender fibers, turbulence modelling, aerodynamic drag, dimensional analysis, data interpolation, stochastic partial differential algebraic equation, numerical simulations, experimental validations
(19 pages, 2009)

173. S. Nickel, M. Schröder, J. Steeg
Planning for home health care services
Keywords: home health care, route planning, meta-heuristics, constraint programming
(23 pages, 2009)

174. G. Dimitroff, A. Szimayer, A. Wagner
Quanto option pricing in the parsimonious Heston model
Keywords: Heston model, multi asset, quanto options, option pricing
(14 pages, 2009) 174. G. Dimitroff, A. Szimayer, A. Wagner

175. S. Herkt, K. Dreßler, R. Pinnau
Model reduction of nonlinear problems in structural mechanics
Keywords: flexible bodies, FEM, nonlinear model reduction, POD
(13 pages, 2009)

176. M. K. Ahmad, S. Didas, J. Iqbal
Using the Sharp Operator for edge detection and nonlinear diffusion
Keywords: maximal function, sharp function,image processing, edge detection, nonlinear diffusion
(17 pages, 2009)

177. M. Speckert, N. Ruf, K. Dreßler, R. Müller, C. Weber, S. Weihe
Ein neuer Ansatz zur Ermittlung von Erprobungslasten für sicherheitsrelevante Bauteile
Keywords: sicherheitsrelevante Bauteile, Kundenbeanspruchung, Festigkeitsverteilung, Ausfallwahrscheinlichkeit, Konfidenz, statistische Unsicherheit, Sicherheitsfaktoren
(16 pages, 2009)

178. J. Jegorovs
Wave based method: new applicability areas
Keywords: Elliptic boundary value problems, inhomogeneous Helmholtz type differential equations in bounded domains, numerical methods, wave based method, uniform B-splines
(10 pages, 2009)

179. H. Lang, M. Arnold
Numerical aspects in the dynamic simulation of geometrically exact rods
Keywords: Kirchhoff and Cosserat rods, geometrically exact rods, deformable bodies, multibody dynamics,artial differential algebraic equations, method of lines, time integration
(21 pages, 2009)

180. H. Lang
Comparison of quaternionic and rotation-free null space formalisms for multibody dynamics
Keywords: Parametrisation of rotations, differential-algebraic equations, multibody dynamics, constrained mechanical systems, Lagrangian mechanics
(40 pages, 2010)

181. S. Nickel, F. Saldanha-da-Gama, H.-P. Ziegler
Stochastic programming approaches for risk aware supply chain network design problems
Keywords: Supply Chain Management, multi-stage stochastic programming, financial decisions, risk
(37 pages, 2010)

182. P. Ruckdeschel, N. Horbenko
Robustness properties of estimators in generalized Pareto Models
Keywords: global robustness, local robustness, finite sample breakdown point, generalized Pareto distribution
(58 pages, 2010)

Status quo: May 2010