Breakdown points and variation exponents of robust M-estimators in linear models

Ivan Mizera*
Comenius University, Bratislava

and

Christine H. Müller
Free University of Berlin

Abstract: The breakdown point behavior of M-estimators in linear models with fixed designs, arising from planned experiments or qualitative factors, is characterized. Particularly, this behavior at fixed designs is quite different from that at designs which can be corrupted by outliers—the situation prevailing in the literature. For fixed designs, the breakdown points of robust M-estimators (those with bounded derivative of the score function), depend on the design and the variation exponent (index) of the score function. This general result implies that the highest breakdown point within all regression equivariant estimators can be attained also by certain M-estimators: those with slowly varying score function, like the Cauchy or slash maximum likelihood estimator. The M-estimators with variation exponent greater than 0, like the L_1 or Huber estimator, exhibit a considerably worse breakdown point behavior. Finally, some known results about the breakdown point of the L_1 estimator are improved and its numerical computation is outlined.

AMS 1991 subject classification: Primary: 62F35, 62F10; Secondary: 62J05, 62J10, 62K99.
Key words and phrases: Breakdown point, L_1 estimator, Linear model, M-estimator, Planned experiments, Regular variation.
Running title: Breakdown points of M-estimators

*Research supported by the Slovak GAS grant 1/1489/94 and by the grant 436 SLK 17/3/95 of the Deutsche Forschungsgemeinschaft

1 Introduction

We consider a *general linear model*

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{y} = (y_1, y_2, \ldots, y_N)^T \in \mathbb{R}^N$ is a *vector of observations*, $\boldsymbol{\beta} \in \mathbb{R}^p$ is an unknown *parameter vector*, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)^T \in \mathbb{R}^N$ a *vector of errors*, and $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)^T \in \mathbb{R}^{N \times p}$ is the known matrix of *design points*, which together constitute a *design*. Let $\widehat{\boldsymbol{\beta}}$ be an estimator of $\boldsymbol{\beta}$. If $\boldsymbol{X}$ is given by an experimentator, it can rightly be assumed that the design points do not contain any gross-errors, outliers and similar phenomena—in other words, they are without errors and, particularly, they are non-stochastic. The same can be said for factors of linear models which are of qualitative nature (in ANOVA models, for instance). In such a context, it is natural to define the *breakdown point* $\varepsilon^*(\widehat{\beta}, \boldsymbol{y}, \boldsymbol{X})$ of $\widehat{\boldsymbol{\beta}}$ as He et al. (1990), Ellis and Morgenthaler (1992), or Müller (1995):

$$\varepsilon^*(\widehat{\boldsymbol{\beta}}, \boldsymbol{y}, \boldsymbol{X}) = \frac{1}{N} \min\{M\colon \sup_{\tilde{\boldsymbol{y}} \in B(\boldsymbol{y}, M)} \|\widehat{\boldsymbol{\beta}}(\tilde{\boldsymbol{y}}, \boldsymbol{X})\| = \infty\},$$

where

$$B(\boldsymbol{y}, M) = \{\tilde{\boldsymbol{y}}\colon\ \mathrm{card}\{n : \tilde{y}_n \neq y_n\} \leq M\}.$$

This approach to the breakdown point (with *fixed design*) differs from that one prevailing in the literature (with *moving design*). The latter assumes that not only the y_i's, but also the design points $\boldsymbol{x}_i$ are vulnerable to errors, and, as a consequence, the definition of the breakdown point allows also for perturbations of the $\boldsymbol{x}_i$'s in $B(\boldsymbol{y}, M)$ (see, for instance, Rousseeuw and Leroy, 1987).

In the present paper, we support the view that for planned experiments and models with qualitative factors, the approach with fixed design is more appropriate. In this vein, an upper bound for the breakdown point was given in Müller (1995): for any regression equivariant estimator (that is, an estimator such that $\widehat{\boldsymbol{\beta}}(\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\vartheta}, \boldsymbol{X}) = \widehat{\boldsymbol{\beta}}(\boldsymbol{y}, \boldsymbol{X}) + \boldsymbol{\vartheta}$ for all $\boldsymbol{y}$, $\boldsymbol{X}$ and $\boldsymbol{\vartheta}$),

$$\varepsilon^*(\widehat{\boldsymbol{\beta}}, \boldsymbol{y}, \boldsymbol{X}) \leq \frac{1}{N} \left\lfloor \frac{N - \mathcal{N}(\boldsymbol{X}) + 1}{2} \right\rfloor, \tag{1}$$

where

$$\mathcal{N}(\boldsymbol{X}) = \max_{\boldsymbol{\beta} \neq \boldsymbol{0}} \mathrm{card}\{n : \boldsymbol{x}_n^T \boldsymbol{\beta} = 0\}$$

is the maximal number of regressors $\boldsymbol{x}_n$ in a subspace of $\mathbb{R}^p$ ($\lfloor u \rfloor$ denotes the largest integer $\leq u$). Note that (1) is the same as the upper bound in the moving design framework, though the latter allows for more general perturbations of the data points—see Rousseeuw and Leroy (1987) for $\mathcal{N}(\boldsymbol{X}) = p - 1$ and Mili and Coakley (1993) for the general case.

It is interesting whether for fixed design the upper bound (1) is attainable and which estimators can achieve it. Müller (1995) showed that the bound (1) can be attained by some trimmed L_p estimators. In the present paper, we address the question whether the bound can be attained also by certain M-estimators.

An *M-estimator* is defined via minimization:

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{y}, \boldsymbol{X}) = \arg\min_{\boldsymbol{\beta}} D(\boldsymbol{\beta}, \boldsymbol{y}, \boldsymbol{X}),$$

where

$$D(\boldsymbol{\beta}, \boldsymbol{y}, \boldsymbol{X}) = \sum_{n=1}^{N} \varphi(y_n - \boldsymbol{x}_n^T \boldsymbol{\beta})$$

is an *objective function* and φ is a given score function from $\mathbb{R}$ to $\mathbb{R}$; in the sequel, we suppose that φ is absolutely continuous, a primitive function of ψ. Note that all M-estimators are regression equivariant.

For the special case of the *location model*— that is, the simplest regression model with $\boldsymbol{x}_1 = \boldsymbol{x}_2 = \cdots = \boldsymbol{x}_N = 1$ and $\boldsymbol{\beta} \in \mathbb{R}^1$—it is known that M-estimators with bounded ψ attain the maximum possible breakdown point of approximately 50% whenever ψ is nondecreasing, corresponding to a convex φ (Huber, 1981, page 54) or φ is unbounded for redescending ψ (Huber, 1984). In regression models with moving design, all M-estimators with nondecreasing ψ have the same low breakdown point $1/N$ as the least squares estimator, as can be shown along the lines of Maronna, Bustos and Yohai (1979). Their finding created an overall impression in the robustness theory that M-estimators possess "bad" breakdown point behavior. However, the result of Maronna, Bustos and Yohai (1979) was based on the convexity of φ and used moving design definition of the breakdown point.

In the setting of fixed design, the whole performance of M-estimators radically changes. As shown below, M-estimators can exhibit high breakdown points, including the highest possible ones. Contrary to the location case, in general linear models the breakdown point of an M-estimator depends on the asymptotic behavior of the function φ, described in terms of the exponent r of regular variation of φ. Together with the design matrix $\boldsymbol{X}$, r determines completely the breakdown behavior of the M-estimator.

The paper is organized as follows. In Section 2 we briefly introduce regularity conditions and state the main result: a dependence of the breakdown point of an M-estimator on its variation exponent r and the design matrix $\boldsymbol{X}$. In Section 3 we discuss the regularity conditions. In Section 4, special attention is paid to two extreme cases: to M-estimators with $r = 0$, like the Cauchy and slash maximum likelihood estimator, which have the highest breakdown point possible for regression equivariant estimators; and to M-estimators with convex φ and $r = 1$, like the L_1, Huber and logistic maximum likelihood estimator, whose performance is much worse. The results for the L_1 estimator improve and complement those of He et al. (1990) and Ellis and Morgenthaler (1992). Section 5 deals with applications: the explicit breakdown points are computed for special linear models and it is shown how choosing an appropriate design, for a given φ, can improve the breakdown point of the M-estimator. Section 6 brings the proofs. Finally, Appendix contains some considerations useful for numerical computation of the breakdown points of M-estimators with $r = 1$.

2 The main result

A measurable function $f : \mathbb{R}^+ \to \mathbb{R}^+$ is called *regularly varying*, if there is a function h such that for all $u > 0$,

$$\lim_{t\to\infty} \frac{f(tu)}{f(t)} = h(u),$$

(see Bingham, Goldie and Teugels, 1987, Chapter 1, or Resnick, 1987, Chapter 0). In such a case, $h(u) = u^r$ for some $r \in \mathbb{R}$, which is called the *exponent* or *index of variation*. If $r = 0$, then f is called *slowly varying*.

We introduce the following assumptions about φ:

(A) Shape: φ is symmetric (even), nondecreasing on $[0, +\infty]$ and nonnegative.

(B) Unboundedness: φ is unbounded.

(C) Subadditivity: there exists $L > 0$ such that

$$\varphi(t + s) \le \varphi(t) + \varphi(s) + L \text{ for all } t, s \ge 0.$$

(D) Regular variation: φ is regularly varying with an exponent $r \ge 0$.

Given $r \ge 0$, we define, using the convention $0^0 = 0$,

$$\mathcal{M}(\boldsymbol{X}, r) = \min\left\{ \operatorname{card} E\colon \sum_{n\in E} |\boldsymbol{x}_n^T\boldsymbol{\beta}|^r \ge \sum_{n\notin E} |\boldsymbol{x}_n^T\boldsymbol{\beta}|^r \text{ for some } \boldsymbol{\beta} \ne \boldsymbol{0} \right\},$$

where E runs over the subsets of $\{1, 2, \ldots, N\}$. For $r = 1$, $\mathcal{M}(\boldsymbol{X}, 1)$ coincides with the minimal possible cardinality of a set $E \subset \{1, 2, \ldots, N\}$ such that

$$\max_{\boldsymbol{\beta}\ne\boldsymbol{0}} \frac{\sum_{n\in E} |\boldsymbol{x}_n^T\boldsymbol{\beta}|}{\sum_{n=1}^N |\boldsymbol{x}_n^T\boldsymbol{\beta}|} \ge \frac{1}{2}.$$

In this form, $\mathcal{M}(\boldsymbol{X}, 1)$ was introduced by He et al. (1990) in the connection with breakdown points of L_1 type estimators. Ellis and Morgenthaler (1992) employed it to derive a lower bound for the exact fit degree of the L_1 estimator. They also pointed out its diagnostic value in the assessment of leverage points in L_1 regression.

Proposition 1. *If $q \geq r \geq 0$ then*

$$\mathcal{M}(\boldsymbol{X}, q) \leq \mathcal{M}(\boldsymbol{X}, r) \leq \mathcal{M}(\boldsymbol{X}, 0) = \left\lfloor \frac{N - \mathcal{N}(\boldsymbol{X}) + 1}{2} \right\rfloor . \tag{2}$$

The following general theorem shows how the breakdown point of an M-estimator depends only on r and $\boldsymbol{X}$, via $\mathcal{M}(\boldsymbol{X}, r)$.

Theorem 1. *If $\widehat{\boldsymbol{\beta}}$ is an M-estimator with φ satisfying (A), (B), (C), and (D) with variation exponent $r \in [0, 1]$, then*

$$\frac{\mathcal{M}(\boldsymbol{X}, r)}{N} \leq \varepsilon^*(\widehat{\boldsymbol{\beta}}, \boldsymbol{y}, \boldsymbol{X}) \leq \frac{\mathcal{M}(\boldsymbol{X}, r) + 1}{N} \tag{3}$$

for all $\boldsymbol{y} \in \mathbb{R}^N$.

The proof of Theorem 1 shows that the upper bound is still true when φ satisfies only (A), (B), and (D) with $r > 1$. Then, in general, the upper bound is not anymore sharp, as the case $r = 2$ shows: the least squares estimator has the breakdown point $1/N$, which is in general not equal to $(\mathcal{M}(\boldsymbol{X}, 2) + 1)/N$. Hence the connection between the breakdown point and variation exponent holds only for robust M-estimators—that is, for those with bounded ψ.

Theorem 1 shows that for the large sample size N, the breakdown point is approximately equal to $\mathcal{M}(\boldsymbol{X}, r)/N$. In special cases, as shown in Section 4, the equality holds: $\varepsilon^*(\widehat{\boldsymbol{\beta}}, \boldsymbol{y}, \boldsymbol{X}) = \mathcal{M}(\boldsymbol{X}, r)/N$.

3 Discussion of the regularity conditions

Concerning shape, it is clear that nonnegativity of φ can be always achieved by adding a constant. The other assumptions in (A) are satisfied by all φ used in practice.

A standard property of regularly varying functions implies that if (D) holds with $r > 0$, then (B) is satisfied. This implication does not hold for $r = 0$; the location case shows that for $r = 0$, the assumption (B) in Theorem 1 is essential.

Since φ and $\varphi + K$ yield the same M-estimates, (C) really means a kind of subadditivity: a constant can be added to φ such that for the resulting φ, $\varphi(t + s) \leq \varphi(t) + \varphi(s)$ holds—not only for $t, s \geq 0$, but for all t, s, as shown by Lemma 1 in Section 6. Subadditivity holds for $\varphi(u) = |u|$; for other functions φ this is not that obvious, nevertheless, (C) can be proved relatively easily from simpler assumptions.

A careful inspection of the proof of Lemma 4.2. in Huber (1984) reveals that if

(i) (A) is fulfilled and $\varphi(0) = 0$;

(ii) there is an $u_0 \le 0$ such that ψ is nondecreasing for $0 < u < u_0$ and nonincreasing for $u_0 < u < \infty$, with $\psi(u) \to 0$ for $u \to \infty$;

then $\eta(t) = \sup_s |\varphi(t+s) - \varphi(s)| - \varphi(t)$ is bounded. This immediately implies (C). (See also Mizera, 1996.) Assumptions (i) and (ii) are satisfied, for instance, by the Cauchy log-likelihood $\varphi(u) = \log(1+u^2)$. Another example is provided by the slash likelihood (for the definition, see Morgenthaler and Tukey, 1991, or Hoaglin, Mosteller and Tukey, 1983).

For convex φ, (C) can be verified by observing that (c denotes the limit of ψ at ∞) for $s, t \ge 0$,

$$\begin{aligned}\varphi(s+t) - \varphi(s) = \int_s^{s+t} \psi(u)\,du \le ct &= \int_0^t \psi(u)\,du + \int_0^t (c - \psi(u))\,du \\ &\le \int_0^t \psi(u)\,du + \int_0^\infty (c - \psi(u))\,du \le \varphi(t) + L\end{aligned}$$

—provided that $\int_0^\infty (c - \psi(u))\,du = L$ is finite. In this way, (C) follows for the Huber as well as for the logistic maximum likelihood estimator (for the definition, see Hampel et al., 1986).

For convex φ, subaddivity also implies boundedness of ψ: just observe that (C) entails

$$\int_s^{s+t} \psi(u)\,du = \varphi(s+t) - \varphi(s) \le \varphi(t) + L.$$

By monotonicity of ψ,

$$\int_s^{s+t} \psi(u)\,du \ge \psi(s)t.$$

Consequently, $\psi(s)t \le \varphi(t) + L$; this cannot be true if $\psi(s) \to \infty$ for $s \to \infty$.

The easiest way to verify (D) is a direct one, using the L'Hospital rule. Also, all functions with a nonzero limit at ∞ are slowly varying; as a consequence, any nondecreasing bounded ψ is slowly varying and this implies, through Proposition 1.5.8 of Bingham, Goldie and Teugels (1987) (see also Resnick, 1987, Theorem 0.6.a), that the corresponding φ is regularly varying with $r = 1$. There are also other criteria for regular variation: for instance, the von Mises property asserts that φ is regularly varying with the exponent r whenever

$$\text{(D1)} \qquad \lim_{u \to \infty} \frac{u\psi(u)}{\varphi(u)} = r.$$

If $u\psi(u)$ is bounded, then (B) immediately implies the slow variation of φ.

If φ is the log-likelihood corresponding to a density f, that is, if $\varphi(u) = -\log f(u)$, then (A) is satisfied whenever f is symmetric and unimodal. Condition (D) in this case reflects the tail behavior of f: it is implied by a stronger condition

(D2) There exist $r \geq 0$, $c > 0$ and $m > 0$ such that

$$\lim_{t\to\infty} \frac{\varphi(t)}{m\,\log(t) + c\,t^r} = 1.$$

Condition (D2) is itself implied by a further stronger condition

(D3) There exist $L^* > 0$, $r \geq 0$, $m \geq 0$ and $c \geq 0$, $m + c > 0$, such that

$$|\varphi(t) - m\,\log(t) - c\,t^r| \leq L^* \text{ for all } t > L^*. \tag{4}$$

A special class of functions satisfying (4) appeared in He et al. (1990): they called those φ satisfying (4) with $m = 0$ and $r = 1$ "score functions of L_1 type". The log-likelihood functions φ satisfying (D2) with $r > 0$ correspond to light-tailed distributions (distributions with exponential tails; for instance, logistic and exponential distribution); the slowly varying φ with $r = 0$ correspond to heavy-tailed distributions (distributions with algebraic tails, like the Student t-distributions with k degrees of freedom, including the Cauchy distribution with $k = 1$). For more background on tail behavior, see Jurečková (1981) or He et al. (1990).

The following proposition summarizes some interrelations among the regularity conditions.

Proposition 2. *Suppose that (A) holds. Then (a) (D1) implies (D); (b) (D2) implies (D) and (B); (c) (D3) implies (D2), and also (C) if $r \in [0, 1]$; (d) (B), (C) and (D) imply $r \in [0, 1]$.*

Summarizing the discussion above, we have (among others) the following examples of estimators satisfying (A), (B), (C) and (D): the L_1, Huber and logistic maximum likelihood estimator ($r = 1$), the Cauchy and slash maximum likelihood estimator ($r = 0$).

4 Special cases: $r = 0$ and $r = 1$

The following theorem is a direct corollary of the upper bound (1) and Proposition 1.

Theorem 2. *For an M-estimator $\widehat{\boldsymbol{\beta}}$ with φ satisfying (A), (B), (C), and (D) with $r = 0$,*

$$\varepsilon^*(\widehat{\boldsymbol{\beta}}, \boldsymbol{y}, \boldsymbol{X}) = \frac{\mathcal{M}(\boldsymbol{X}, 0)}{N} = \frac{1}{N}\left\lfloor \frac{N - \mathcal{N}(\boldsymbol{X}) + 1}{2} \right\rfloor$$

holds for all $\boldsymbol{y} \in \mathbb{R}^N$ and $\boldsymbol{X}$.

In other words, M-estimators with slowly varying φ, like the Cauchy or slash maximum likelihood estimator, have the highest breakdown point possible for regression equivariant estimators.

¿From Proposition 1 follows that this does not hold, in general, for M-estimators with $r = 1$ (see also Section 5). For certain special cases of these estimators, a sharpening of Theorem 1, akin to Theorem 2, can be proved. We prove this fact for the L_1 estimator.

Theorem 3. *For the L_1 estimator $\widehat{\boldsymbol{\beta}}_1$,*

$$\varepsilon^*(\widehat{\boldsymbol{\beta}}_1, \boldsymbol{y}, \boldsymbol{X}) = \frac{\mathcal{M}(\boldsymbol{X}, 1)}{N} \tag{5}$$

holds for all $\boldsymbol{y} \in \mathbb{R}^N$ and $\boldsymbol{X}$.

Proving Theorem 3, we have only to show that, in view of (3),

$$\varepsilon^*(\widehat{\boldsymbol{\beta}_1}, \boldsymbol{y}, \boldsymbol{X}) \leq \frac{\mathcal{M}(\boldsymbol{X}, 1)}{N}. \tag{6}$$

We first establish (6) for $\boldsymbol{y} = \mathbf{0}$. To finish the proof of Theorem 3, it is sufficient to verify that the L_1 estimator is of uniform variation. (See Section 6 for the details.)

We say that an estimator $\widehat{\boldsymbol{\beta}}$ is of *uniform variation*, if, given $\boldsymbol{X}$, for every $\delta > 0$ there is a $K > 0$ such that whenever $\|\boldsymbol{y}^1 - \boldsymbol{y}^2\| \leq \delta$, then for any $\boldsymbol{\beta}^1 \in \widehat{\boldsymbol{\beta}}(\boldsymbol{y}^1, \boldsymbol{X})$ there is $\boldsymbol{\beta}^2 \in \widehat{\boldsymbol{\beta}}(\boldsymbol{y}^2, \boldsymbol{X})$ such that $\|\boldsymbol{\beta}^1 - \boldsymbol{\beta}^2\| \leq K$. Uniform variation implies—as shown by Lemma 4—that the breakdown point $\varepsilon^*(\widehat{\boldsymbol{\beta}_1}, \boldsymbol{y}, \boldsymbol{X})$ does not exceed the breakdown point at $\mathbf{0}$. Ellis and Morgenthaler (1992) used a similar idea for the transition from the upper bound on the exactfit point to the upper bound on the breakdown point for an arbitrary $\boldsymbol{y}$. Instead of uniform variation they used uniform continuity (which is more stringent and implies uniform variation).

It is plausible that the analogue of Theorem 3 holds for some other L_1-type estimators: particularly, we were able to prove (6) with $\boldsymbol{y} = \mathbf{0}$ for the Huber estimator. We conjecture that this estimator is also of uniform variation, and, as a consequence, satisfies (5). In this connection, Lemma 5.3 of He et al. (1990) should be quoted: one of its possible interpretations (its original statement is in a slightly vague form) would yield uniform variation of all L_1-type estimators. Unfortunately, that version which could be proved—matching the needs of the original paper (He and Portnoy, personal communication)—seems not to be helpful for our purposes.

5 Applications

For designs with minimum support, like the classical A- and D- optimal designs for polynomial regression and the designs for the one-way layout, the breakdown points of all M-estimators with φ satisfying (A), (B), (C) and (D) with $r \in [0,1]$ are equal and attain the maximum value if the numbers of repetitions are equal. This is a consequence of the following proposition.

Proposition 3. *Suppose that* $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \{\boldsymbol{\chi}_1, \ldots, \boldsymbol{\chi}_p\} \subset \mathbb{R}^p$. *Let* $N_i = \sum_{n=1}^N 1_{\{\boldsymbol{\chi}_i\}}(x_n)$ *for* $i = 1, \ldots, p$. *If* $\boldsymbol{\chi}_1, \ldots, \boldsymbol{\chi}_p$ *are linearly independent, then*

$$\mathcal{M}(\boldsymbol{X}, r) = \left\lfloor \frac{N - \mathcal{N}(\boldsymbol{X}) + 1}{2} \right\rfloor = \left\lfloor \frac{1 + \min_{i=1,\ldots,p} N_i}{2} \right\rfloor \leq \left\lfloor \frac{N+p}{2p} \right\rfloor$$

for every $r \in [0,1]$.

In the setting of Proposition 3, or if $r = 0$, the quantity $\mathcal{M}(\boldsymbol{X}, r)$ can be easily computed, since $\mathcal{N}(\boldsymbol{X})$ can be (see Müller, 1995, 1996a, 1996b, for a numerical algorithm Mili and Coakley, 1993). In other cases, computing $\mathcal{M}(\boldsymbol{X}, r)$ requires special considerations.

In the sequel we concentrate on the important case $r = 1$, which provides also lower bounds for every $r \in (0,1)$, in view of Proposition 1. For $r = 1$, we can construct an algorithm to compute $\mathcal{M}(\boldsymbol{X}, 1)$ numerically, based on certain convexity and linearity properties which hold in this case; see Appendix for more details. But, for special linear models, we are also able to give explicit formulas.

For the linear regression through the origin, we have the following obvious result (see also He et al., 1990).

Proposition 4. *If, for* $n = 1, \ldots, N$, $\boldsymbol{x}_n = t_n \in \mathbb{R}$ *with* $|t_1| \leq |t_2| \leq \cdots \leq |t_N|$ *then*

$$\mathcal{M}(\boldsymbol{X}, 1) = \min \left\{ M\colon \sum_{n=1}^{N-M} |t_n| \leq \sum_{n=N-M+1}^{N} |t_n| \right\} \leq \left\lfloor \frac{N+1}{2} \right\rfloor.$$

The equality holds if and only if $|t_1| = |t_2| = \cdots = |t_N|$.

We also can write an explicit formula for $\mathcal{M}(\boldsymbol{X}, 1)$ in the univariate linear regression with a constant term.

Proposition 5. *If, for* $n = 1, \ldots, N$, $\boldsymbol{x}_n = (1, t_n)^T \in \mathbb{R}^2$ *with* $t_1 \leq t_2 \leq \cdots \leq t_N$, *then*

$$\begin{aligned} \mathcal{M}(\boldsymbol{X}, 1) = \min \Bigg\{ M\colon & \sum_{n=1}^{N-M} (t_n - t_1) \leq \sum_{n=N-M+1}^{N} (t_n - t_1) \\ & \textit{or} \sum_{n=M+1}^{N} (t_N - t_n) \leq \sum_{n=1}^{M} (t_N - t_n) \Bigg\} \\ \leq \left\lfloor \frac{N+3}{4} \right\rfloor. \end{aligned}$$

Proposition 5 shows that the upper bound of $N/4$ is attained by designs with equispaced design points t_n, as well as by the classical D-optimal design which puts a half of the observation at the each endpoint of the design region. There are also other designs attaining the upper bound — for instance, the design with L observations at 1 and 3 and $2L$ observations at 2.

Moreover, note that in the linear regression setting of Proposition 5, M-estimators satisfying (D) with $r = 0$ have the breakdown point $\lfloor N/2 \rfloor$ if the design points $\boldsymbol{x}_n$ are pairwise different (see Müller, 1995). That means that the Cauchy estimator (among others) has a much higher breakdown point than the L_1 estimator.

Proposition 5 gives also upper bounds for the breakdown point in other linear models.

Proposition 6. *Let, for $n = 1, \ldots, N$, $\boldsymbol{x}_n = (x_n^1, \ldots, x_n^p)^T$ with $x_1^i \le x_2^i \le \cdots \le x_N^i$ for $i \in \{1, \ldots, p\}$. If there exists $\boldsymbol{A} \in \mathbb{R}^{2\times p}$ such that $\boldsymbol{A}\boldsymbol{x}_n = (1, x_n^i)^T$ for $n = 1, \ldots, N$, then*

$$\begin{aligned} \mathcal{M}(\boldsymbol{X}, 1) \le \min \Bigg\{ M\colon & \sum_{n=1}^{N-M} (x_n^i - x_1^i) \le \sum_{n=N-M+1}^{N} (x_n^i - x_1^i) \\ & \textit{or} \sum_{n=M+1}^{N} (x_N^i - x_n^i) \le \sum_{n=1}^{M} (x_N^i - x_n^i) \Bigg\} \\ \le \left\lfloor \frac{N+3}{4} \right\rfloor . \end{aligned}$$

Proposition 6 provides, in particular for polynomial regression and for multiple linear regression, an upper bound for the breakdown point of the L_1 and Huber estimators, as well as a complete proof of the Ellis' and Morgenthaler's upper bound of 1/4. Ellis and Morgenthaler (1992) showed, for lattice designs, that there are designs attaining this upper bound.

6 Proofs

Proof of Proposition 1. We first prove the inequality part of (2). Take an arbitrary $\boldsymbol{\beta} \neq 0$ and $E \subset \{1, 2, \ldots, N\}$ with card $E = \mathcal{M}(\boldsymbol{X}, r)$ and

$$\sum_{n\in E} |\boldsymbol{x}_n^T \boldsymbol{\beta}|^r - \sum_{n\notin E} |\boldsymbol{x}_n^T \boldsymbol{\beta}|^r \ge 0. \tag{7}$$

If there are $i \in E$ and $j \notin E$ such that $|\boldsymbol{x}_i^T\boldsymbol{\beta}| < |\boldsymbol{x}_j^T\boldsymbol{\beta}|$, then we can exchange i and j—that is, we can consider $(E \cup \{j\}) \setminus \{i\}$ instead of E. Thus, without loss of generality we can assume that

$$d = \min\{|\boldsymbol{x}_n^T\boldsymbol{\beta}|\colon n \in E\} \ge \max\{|\boldsymbol{x}_n^T\boldsymbol{\beta}|\colon n \notin E\}.$$

Then

$$\begin{aligned}0 &\le d^{q-r}\left(\sum_{n\in E}|\boldsymbol{x}_n^T\boldsymbol{\beta}|^r - \sum_{n\notin E}|\boldsymbol{x}_n^T\boldsymbol{\beta}|^r\right)\\ &\le \sum_{n\in E}|\boldsymbol{x}_n^T\boldsymbol{\beta}|^r|\boldsymbol{x}_n^T\boldsymbol{\beta}|^{q-r} - \sum_{n\notin E}|\boldsymbol{x}_n^T\boldsymbol{\beta}|^r|\boldsymbol{x}_n^T\boldsymbol{\beta}|^{q-r},\end{aligned}$$

hence $\mathcal{M}(\boldsymbol{X},q)\le\mathcal{M}(\boldsymbol{X},r)$.

Now we show the equality in (2). By definition of $\mathcal{N}(\boldsymbol{X})$, there exists $\boldsymbol{\beta}\neq 0$ and $E_0\subset\{1,2,\dots,N\}$ with $\operatorname{card}E_0=\mathcal{N}(\boldsymbol{X})$ such that $|\boldsymbol{x}_n^T\boldsymbol{\beta}|^0=0$ for all $n\in E_0$ and $|\boldsymbol{x}_n^T\boldsymbol{\beta}|^0=1$ for all $n\notin E_0$. Then for any subset $E\subset\{1,2,\dots,N\}\setminus E_0$ with $\operatorname{card}E=\lfloor(N-\mathcal{N}(\boldsymbol{X})+1)/2\rfloor$ we have

$$\begin{aligned}\sum_{n\in E}|\boldsymbol{x}_n^T\boldsymbol{\beta}|^0 &= \left\lfloor\tfrac{1}{2}(N-\mathcal{N}(\boldsymbol{X})+1)\right\rfloor\\ &\ge N-\left\lfloor\tfrac{1}{2}(N-\mathcal{N}(\boldsymbol{X})+1)\right\rfloor-\mathcal{N}(\boldsymbol{X})=\sum_{n\notin E}|\boldsymbol{x}_n^T\boldsymbol{\beta}|^0;\end{aligned}$$

hence $\mathcal{M}(\boldsymbol{X},0)\le\lfloor(N-\mathcal{N}(\boldsymbol{X})+1)/2\rfloor$. To see the reverse inequality, take any $\boldsymbol{\beta}\neq 0$ and $E\subset\{1,2,\dots,N\}$ with $\operatorname{card}E=\mathcal{M}(\boldsymbol{X},0)$ such that (7) holds with $r=0$. Since at most $\mathcal{N}(\boldsymbol{X})$ of the $\boldsymbol{x}_n^T\boldsymbol{\beta}$'s satisfy $|\boldsymbol{x}_n^T\boldsymbol{\beta}|^0=0$, we have

$$\begin{aligned}\mathcal{M}(\boldsymbol{X},0)\ge\sum_{n\in E}|\boldsymbol{x}_n^T\boldsymbol{\beta}|^0 &\ge \sum_{n\notin E}|\boldsymbol{x}_n^T\boldsymbol{\beta}|^0\\ &\ge N-\mathcal{M}(\boldsymbol{X},0)-\mathcal{N}(\boldsymbol{X}),\end{aligned}$$

and this implies that $2\mathcal{M}(\boldsymbol{X},0)\ge N-\mathcal{N}(\boldsymbol{X})$; that is, $\mathcal{M}(\boldsymbol{X},0)\ge\lfloor(N-\mathcal{N}(\boldsymbol{X})+1)/2\rfloor$. $\square$

Before proving Theorem 1, we state three lemmas.

Lemma 1. *Conditions (A) and (C) imply*

$$\varphi(t+s)\le\varphi(t)+\varphi(s)+L \text{ for all } t,s\in\mathbb{R}.$$

Proof. We prove the equivalent assertion: for any $s,t\in\mathbb{R}$, $\varphi(s-t)\ge\varphi(s)-\varphi(t)-L$. If $0\le t<s$, then $\varphi(s)=\varphi(s-t+t)\le\varphi(s-t)+\varphi(t)+L$. If $s<t\le 0$, then, due to the symmetry of φ, we have $\varphi(s)=\varphi(-s)=\varphi(-(s-t)-t)\le\varphi(s-t)+\varphi(t)+L$. Finally, if $|s|\le|t|$, then $\varphi(s-t)-\varphi(s)\ge 0-\varphi(t)-L$, since φ is nonnegative and nondecreasing. $\square$

Lemma 2. *Suppose that φ satisfies conditions (A) and (D) and let $t_k\to\infty$ and $u_k\to u$ be arbitrary. Then*

$$\lim_{k\to\infty}\frac{\varphi(u_kt_k)}{\varphi(t_k)}=u^r,$$

for $k\to\infty$, whenever (a) $r>0$, $u_k\ge 0$ and $u\ge 0$; or (b) $r=0$ and $u>0$

Proof. The lemma follows from the standard property of regularly varying functions: see, for instance, Bingham, Goldie and Teugels (1987, Theorem 1.5.7) or Resnick (1987, Proposition 0.5). □

Lemma 3. *If* $\mathcal{N}(\boldsymbol{X}) < N$*, then for all* $r \geq 0$ *there is a* $\boldsymbol{\beta} \neq \mathbf{0}$ *and a subset* $E \subseteq \{1, 2, \ldots, N\}$ *with* $\operatorname{card} E = \mathcal{M}(\boldsymbol{X}, r) + 1$ *such that*

$$\sum_{n \in E} |\boldsymbol{x}_n^T \boldsymbol{\beta}|^r > \sum_{n \notin E} |\boldsymbol{x}_n^T \boldsymbol{\beta}|^r.$$

Proof. Let $\boldsymbol{\beta} \neq 0$ and $E \subset \{1, 2, \ldots, N\}$ with $\operatorname{card} E = \mathcal{M}(\boldsymbol{X}, r)$ so that

$$\sum_{n \in E} |\boldsymbol{x}_n^T \boldsymbol{\beta}|^r \geq \sum_{n \notin E} |\boldsymbol{x}_n^T \boldsymbol{\beta}|^r \tag{8}$$

is satisfied. If the inequality in (8) is sharp then it remains sharp when we add an element to E. If we have equality in (8) and there exists $n \notin E$ with $|\boldsymbol{x}_n^T \boldsymbol{\beta}| \neq 0$ then we can add this n to E so that for $E \cup \{n\}$ the inequality in (8) is sharp. If $|\boldsymbol{x}_n^T \boldsymbol{\beta}| = 0$ for all $n \notin E$ then equality in (8) implies $|\boldsymbol{x}_n^T \boldsymbol{\beta}| = 0$ for all $n = 1, 2, \ldots, N$ so that $\mathcal{N}(\boldsymbol{X}) = N$. This contradicts the assumptions. □

Proof of Theorem 1. *Lower bound.* Due to the regression equivariance of M-estimators, we may assume that $\widehat{\boldsymbol{\beta}}(\boldsymbol{y}, \boldsymbol{X}) = \mathbf{0}$, without loss of generality. Let $M = N\varepsilon^*(\widehat{\boldsymbol{\beta}}, \boldsymbol{y}, \boldsymbol{X})$. Then for every $k \in \mathbb{N}$ there is a $\boldsymbol{y}^k \in B(\boldsymbol{y}, M)$, such that $\|\widehat{\boldsymbol{\beta}}(\boldsymbol{y}^k, \boldsymbol{X})\| > k$. Again without restricting generality we may suppose that $y_n^k = y_n$ for $n = 1, 2, \ldots, N - M$. For brevity, $\widehat{\boldsymbol{\beta}}(\boldsymbol{y}^k, \boldsymbol{X})$ is denoted by $\boldsymbol{\beta}^k$. We have that $\|\boldsymbol{\beta}^k\| \to \infty$ and, in view of the compactness of the unit sphere in $\mathbb{R}^p$, we may also suppose that

$$\frac{\boldsymbol{\beta}^k}{\|\boldsymbol{\beta}^k\|} \to \boldsymbol{\beta}^0,$$

passing otherwise to a subsequence. Since

$$\begin{aligned} 0 &\geq D(\boldsymbol{\beta}^k, \boldsymbol{y}^k, \boldsymbol{X}) - D(\mathbf{0}, \boldsymbol{y}^k, \boldsymbol{X}) \\ &= \sum_{n=1}^{N-M} \varphi(y_n - \boldsymbol{x}_n^T \boldsymbol{\beta}^k) + \sum_{n=N-M+1}^{N} \varphi(y_n^k - \boldsymbol{x}_n^T \boldsymbol{\beta}^k) \\ &\qquad - \sum_{n=1}^{N-M} \varphi(y_n) - \sum_{n=N-M+1}^{N} \varphi(y_n^k), \end{aligned}$$

we have by Lemma 1 and the symmetry of φ

$$0 \geq \sum_{n=1}^{N-M} \varphi(\boldsymbol{x}_n^T \boldsymbol{\beta}^k) - 2 \sum_{n=1}^{N-M} \varphi(y_n) - \sum_{n=N-M+1}^{N} \varphi(\boldsymbol{x}_n^T \boldsymbol{\beta}^k) - NL. \tag{9}$$

Dividing (9) by $\varphi(\|\boldsymbol{\beta}^k\|)$ and letting $k \to \infty$, we obtain

$$\begin{aligned} 0 \geq \sum_{n=1}^{N-M} \liminf_{k\to\infty} & \frac{\varphi\left(\|\boldsymbol{\beta}^k\| \left|\boldsymbol{x}_n^T \frac{\boldsymbol{\beta}^k}{\|\boldsymbol{\beta}^k\|}\right|\right)}{\varphi(\|\boldsymbol{\beta}^k\|)} \\ & - \sum_{n=N-M+1}^{N} \limsup_{k\to\infty} \frac{\varphi\left(\|\boldsymbol{\beta}^k\| \left|\boldsymbol{x}_n^T \frac{\boldsymbol{\beta}^k}{\|\boldsymbol{\beta}^k\|}\right|\right)}{\varphi(\|\boldsymbol{\beta}^k\|)}. \end{aligned} \tag{10}$$

If $r \in (0,1]$ then, in view of Lemma 2(a), the inequality (10) acquires the form

$$0 \geq \sum_{n=1}^{N-M} |\boldsymbol{x}_n^T \boldsymbol{\beta}^0|^r - \sum_{n=N-M+1}^{N} |\boldsymbol{x}_n^T \boldsymbol{\beta}^0|^r,$$

which implies that $M \geq \mathcal{M}(\boldsymbol{X}, r)$. Consequently, we have $\varepsilon^*(\widehat{\boldsymbol{\beta}}, \boldsymbol{y}, \boldsymbol{X}) \geq \mathcal{M}(\boldsymbol{X}, r)/N$ in this case.

Suppose now that $r = 0$. If $t_k \to \infty$ and $u_k \to 0$, then

$$0 \leq \liminf_{k\to\infty} \frac{\varphi(u_k t_k)}{\varphi(t_k)} \leq \limsup_{k\to\infty} \frac{\varphi(u_k t_k)}{\varphi(t_k)} \leq \lim_{k\to\infty} \frac{\varphi(t_k)}{\varphi(t_k)} = 1, \tag{11}$$

by (A). Since $\boldsymbol{x}_n^T \boldsymbol{\beta}^0 = 0$ for at most $\mathcal{N}(\boldsymbol{X})$ of the $\boldsymbol{x}_n$'s, (10), (11) and Lemma 2(b) entail

$$0 \geq N - M - \mathcal{N}(\boldsymbol{X}) - M,$$

hence $M \geq (N - \mathcal{N}(\boldsymbol{X}))/2$, and consequently, in view of Proposition 1,

$$\varepsilon^*(\widehat{\boldsymbol{\beta}}, \boldsymbol{y}, \boldsymbol{X}) \geq \frac{1}{N} \left\lfloor \frac{N - \mathcal{N}(\boldsymbol{X}) + 1}{2} \right\rfloor = \frac{\mathcal{M}(\boldsymbol{X}, 0)}{N}.$$

Upper bound. We consider the case $r = 0$ first. Here we see at once that the upper bound (1) and Proposition 1 yield even that

$$\varepsilon^*(\widehat{\boldsymbol{\beta}}, \boldsymbol{y}, \boldsymbol{X}) \leq \frac{1}{N} \left\lfloor \frac{N - \mathcal{N}(\boldsymbol{X}) + 1}{2} \right\rfloor = \frac{\mathcal{M}(\boldsymbol{X}, 0)}{N}.$$

Now, let $r \in (0,1]$. Suppose that $M = \mathcal{M}(\boldsymbol{X}, r) + 1$. If $\mathcal{N}(\boldsymbol{X}) = N$, the upper bound (1) shows that there is nothing to prove; hence suppose that $\mathcal{N}(\boldsymbol{X}) < N$. By Lemma 3, there is a set $E \subseteq \{1, 2, \ldots, N\}$, with $\operatorname{card} E = M$, and $\boldsymbol{\beta}^1 \neq 0$ such that

$$\sum_{n\in E} |\boldsymbol{x}_n^T \boldsymbol{\beta}^1|^r > \sum_{n\notin E} |\boldsymbol{x}_n^T \boldsymbol{\beta}^1|^r. \tag{12}$$

Without restriction of generality we may now assume that $E = \{N - M + 1, N - M + 2, \ldots, N\}$. Let $\boldsymbol{\beta}^k = k\boldsymbol{\beta}^1$. Define $\boldsymbol{y}^k$ to satisfy $y_n^k = y_n$ for $n = 1, 2, \ldots, N - M$ and $y_n^k = \boldsymbol{x}_n^T \boldsymbol{\beta}^k$ for $n \in E$.

Suppose that there is a compact $\mathcal{K}$ such that $\tilde{\boldsymbol{\beta}}^k = \widehat{\boldsymbol{\beta}}(\boldsymbol{y}^k, \boldsymbol{X}) \in \mathcal{K}$ for all k. Then,

$$\begin{aligned} 0 < d_k &= D(\boldsymbol{\beta}^k, \boldsymbol{y}^k, \boldsymbol{X}) - D(\tilde{\boldsymbol{\beta}}^k, \boldsymbol{y}^k, \boldsymbol{X}) \\ &= \sum_{n=1}^{N-M} \varphi(y_n - \boldsymbol{x}_n^T \boldsymbol{\beta}^k) - \sum_{n=1}^{N-M} \varphi(y_n - \boldsymbol{x}_n^T \tilde{\boldsymbol{\beta}}^k) \\ &+ \sum_{n=N-M+1}^{N} \varphi(0) - \sum_{n=N-M+1}^{N} \varphi(\boldsymbol{x}_n^T \boldsymbol{\beta}^k - \boldsymbol{x}_n^T \tilde{\boldsymbol{\beta}}^k). \end{aligned}$$

Since $\tilde{\boldsymbol{\beta}}^k$, and consequently $\boldsymbol{x}_n^T \tilde{\boldsymbol{\beta}}^k$ are in a compact set, dividing by $\varphi(k)$ and letting $k \to \infty$ yields, in view of (A) and Lemma 2(a), that

$$0 \le \lim_{k\to\infty} \frac{d_k}{\varphi(k)}$$

$$= \sum_{n=1}^{N-M} \lim_{k\to\infty} \frac{\varphi\left(k\left(\boldsymbol{x}_n^T \boldsymbol{\beta}^1 - \frac{1}{k} y_n\right)\right)}{\varphi(k)} - \sum_{n=N-M+1}^{N} \lim_{k\to\infty} \frac{\varphi(k(\boldsymbol{x}_n^T \boldsymbol{\beta}^1 - \frac{1}{k} \boldsymbol{x}_n^T \tilde{\boldsymbol{\beta}}^k))}{\varphi(k)}$$

$$= \sum_{n=1}^{N-M} |\boldsymbol{x}_n^T \boldsymbol{\beta}^1|^r - \sum_{n=N-M+1}^{N} |\boldsymbol{x}_n^T \boldsymbol{\beta}^1|^r.$$

This is in contradiction with (12); hence, $\varepsilon^*(\widehat{\boldsymbol{\beta}}, \boldsymbol{y}, \boldsymbol{X}) \le M/N$. □

Proof of Proposition 2. For (a), see Resnick (1987, Proposition 0.7.a). Concerning (b): it is clear that (D2) implies (B). (D2) with $r > 0$ implies

$$\lim_{t\to\infty} \frac{\varphi(t)}{c\, t^r} = 1,$$

thus for every $u > 0$ we have

$$\lim_{t\to\infty} \frac{\varphi(u\,t)}{\varphi(t)} = \lim_{t\to\infty} \frac{\varphi(u\,t)}{c\,(u\,t)^r}\, u^r\, \frac{c\,t^r}{\varphi(t)} = u^r$$

and hence (D) holds. (D2) with $r = 0$ implies

$$\lim_{t\to\infty} \frac{\varphi(t)}{m \log(t)} = 1,$$

hence for every $u > 0$,

$$\lim_{t\to\infty} \frac{\varphi(u\,t)}{\varphi(t)} = \lim_{t\to\infty} \frac{\varphi(u\,t)}{m \log(u\,t)} \frac{m \log(u) + m \log(t)}{\varphi(t)} = 1$$

and (D) holds as well. Statement (c) is obvious. Finally, (C) yields that

$$\frac{\varphi(2t)}{\varphi(t)} \le 2 + \frac{L}{\varphi(t)}.$$

Passing to limit on both sides and using (D) and (B), we obtain that $2^r \le 2$; hence $r \le 1$, proving (d). □

The following sequence of lemmas proves Theorem 3.

Lemma 4. *Suppose that $\widehat{\boldsymbol{\beta}}$ is of uniform variation. Then for all $\boldsymbol{y} \in \mathbb{R}^N$,*

$$\varepsilon^*(\widehat{\boldsymbol{\beta}}, \boldsymbol{y}, \boldsymbol{X}) \leq \varepsilon^*(\widehat{\boldsymbol{\beta}}, \mathbf{0}, \boldsymbol{X}).$$

Proof. Let $M = N\varepsilon^*(\widehat{\boldsymbol{\beta}}, \mathbf{0}, \boldsymbol{X})$; then there is a sequence $\tilde{\boldsymbol{y}}^k$ such that (without restricting generality) $\tilde{y}_n^k = 0$ for $n = 1, 2, \ldots, N - M$, and

$$\|\widehat{\boldsymbol{\beta}}(\tilde{\boldsymbol{y}}^k, \boldsymbol{X})\| \to \infty. \tag{13}$$

Given $\boldsymbol{y}$, let $\delta = \|\boldsymbol{y} - \mathbf{0}\|$. Let $\boldsymbol{y}^k$ be the vector with components $y_n^k = y_n + \tilde{y}_n^k$ for $n = 1, 2, \ldots, N$. Since

$$\|\boldsymbol{y}^k - \tilde{\boldsymbol{y}}^k\| = \sum_{n=1}^{N} y_n^2 = \|\boldsymbol{y} - \mathbf{0}\| = \delta,$$

the uniform variation of $\boldsymbol{\beta}$ yields that there is $K > 0$ and $\widehat{\boldsymbol{\beta}}(\boldsymbol{y}^k, \boldsymbol{X})$ such that

$$\|\widehat{\boldsymbol{\beta}}(\boldsymbol{y}^k, \boldsymbol{X}) - \widehat{\boldsymbol{\beta}}(\tilde{\boldsymbol{y}}^k, \boldsymbol{X})\| \leq K$$

for all k. Thus, $\|\widehat{\boldsymbol{\beta}}(\boldsymbol{y}^k, \boldsymbol{X})\| \to \infty$ by (13), so that $\varepsilon^*(\widehat{\boldsymbol{\beta}}, \boldsymbol{y}, \boldsymbol{X}) \leq M/N$, proving the lemma. □

Lemma 5. *For any $\boldsymbol{X}$, $\varepsilon^*(\widehat{\boldsymbol{\beta}_1}, \mathbf{0}, \boldsymbol{X}) \leq \mathcal{M}(\boldsymbol{X}, 1)/N$.*

Proof. Let $M = \mathcal{M}(\boldsymbol{X}, 1)$. By the definition of $\mathcal{M}(\boldsymbol{X}, 1)$, there is a $\boldsymbol{\beta}^1 \neq \mathbf{0}$ and E with card $E = M$ such that for $\boldsymbol{\beta} = \boldsymbol{\beta}^1$,

$$\sum_{n \in E} |\boldsymbol{x}_n^T \boldsymbol{\beta}| \geq \sum_{n \notin E} |\boldsymbol{x}_n^T \boldsymbol{\beta}|. \tag{14}$$

Then (14) holds, with the same E, also for $\boldsymbol{\beta} = \boldsymbol{\beta}^k = k\boldsymbol{\beta}^1$, for all $k \in \mathbb{N}$. Without restriction of generality we may again suppose that $E = \{N - M + 1, N - M + 2, \ldots, N\}$.

For $k \in \mathbb{N}$, let $y_n^k = 0$ for $n = 1, 2, \ldots, N - M$ and $y_n^k = \boldsymbol{x}_n^T \boldsymbol{\beta}^k$ otherwise. For the L_1 estimator, $\varphi(u) = \varphi_1(u) = |u|$; by (14),

$$\begin{aligned} &D(\boldsymbol{\beta}^k, \boldsymbol{y}^k, \boldsymbol{X}) - D(\mathbf{0}, \boldsymbol{y}^k, \boldsymbol{X}) \\ &\qquad = \sum_{n=1}^{N-M} |\boldsymbol{x}_n^T \boldsymbol{\beta}^k| - \sum_{n=N-M+1}^{N} |\boldsymbol{x}_n^T \boldsymbol{\beta}^k| \leq 0. \end{aligned}$$

By the convexity of the argmin set, there is a $\tilde{\boldsymbol{\beta}}^1$ minimizing $D(\boldsymbol{\beta}, \boldsymbol{y}^1, \boldsymbol{X})$ which is equal neither to $\mathbf{0}$ nor to $\boldsymbol{\beta}^1$. It is easily seen, directly from the definition, that the L_1 estimator is scale equivariant: $\widehat{\boldsymbol{\beta}}(s\boldsymbol{y}, \boldsymbol{X}) = s\widehat{\boldsymbol{\beta}}(\boldsymbol{y}, \boldsymbol{X})$. Therefore, $\tilde{\boldsymbol{\beta}}^k = k\tilde{\boldsymbol{\beta}}^1$ minimizes $D(\boldsymbol{\beta}, \boldsymbol{y}^k, \boldsymbol{X})$. Since $\|\tilde{\boldsymbol{\beta}}^k\| \to \infty$, the lemma follows. □

The proof of the uniform variation of the L_1 estimator uses the idea of directional derivatives, borrowed from the proof of Lemma 5.3 of He et al. (1990). Define

$$\begin{aligned} S_{\boldsymbol{b}}(\alpha) &= \sum_{n=1}^{N} |y_n - \boldsymbol{x}_n^T(\boldsymbol{\beta} + \alpha \boldsymbol{b})|, \\ R &= \{n \colon y_n = \boldsymbol{x}_n^T \boldsymbol{\beta}\} \end{aligned}$$

For the directional derivative

$$S'_{\boldsymbol{b}}(0) = \lim_{\alpha \downarrow 0} \frac{S_{\boldsymbol{b}}(\alpha) - S_{\boldsymbol{b}}(0)}{\alpha},$$

we have

$$\begin{aligned} &S'_{\boldsymbol{b}}(0) = -\sum_{n \notin R} \operatorname{sign}(y_n - \boldsymbol{x}_n^T \boldsymbol{\beta})\, \boldsymbol{x}_n^T \boldsymbol{b} + \sum_{n \in R} |\boldsymbol{x}_n^T \boldsymbol{b}|, \\ &S'_{\boldsymbol{b}}(0)\, \alpha \le S_{\boldsymbol{b}}(\alpha) - S_{\boldsymbol{b}}(0), \end{aligned}$$

and $S'_{\boldsymbol{b}}(0) > 0$ whenever α_0 exists such that $S_{\boldsymbol{b}}(\alpha) > S_{\boldsymbol{b}}(0)$ for all $0 < \alpha < \alpha_0$. Define, analogously to He et al. (1990), page 1208,

$$\begin{aligned} \mathcal{S}(\boldsymbol{b}, R, \boldsymbol{X}) &= \{\boldsymbol{s} = (s_n)_{n \notin R} \in \{-1, 1\}^{N - \operatorname{card} R} \colon \\ &\qquad\qquad -\sum_{n \notin R} s_n\, \boldsymbol{x}_n^T \boldsymbol{b} + \sum_{n \in R} |\boldsymbol{x}_n^T \boldsymbol{b}| > 0\}, \\ g(\boldsymbol{b}, R, \boldsymbol{X}) &= \min_{\boldsymbol{s} \in \mathcal{S}(\boldsymbol{b}, R, \boldsymbol{X})} \{-\sum_{n \notin R} s_n\, \boldsymbol{x}_n^T \boldsymbol{b} + \sum_{n \in R} |\boldsymbol{x}_n^T \boldsymbol{b}|\}, \\ c(R, \boldsymbol{X}) &= \min\{g(\boldsymbol{b}, R, \boldsymbol{X}) \colon \|\boldsymbol{b}\| = 1\}, \\ c(\boldsymbol{X}) &= \min\{c(R, \boldsymbol{X}) \colon R \subset \{1, 2, \ldots, N\}\}. \end{aligned}$$

We have that $g(\boldsymbol{b}, R, \boldsymbol{X}) > 0$, $c(R, \boldsymbol{X}) > 0$, $c(\boldsymbol{X}) > 0$.

Lemma 6. *The L_1 estimator is of uniform variation.*

Proof. Fix $\delta > 0$ and $\boldsymbol{y} = (y_1, y_2, \ldots, y_N)^T \in \mathbb{R}^N$. Take any $\boldsymbol{y}^0 \in \mathbb{R}^N$ with $|y_n^0 - y_n| \le \delta$ for all $n = 1, 2, \ldots, N$ and any $\boldsymbol{\beta}^0 \in \widehat{\boldsymbol{\beta}}_1(\boldsymbol{y}^0, \boldsymbol{X})$. We shall prove that there exists $\boldsymbol{\beta} \in \widehat{\boldsymbol{\beta}}_1(\boldsymbol{y}, \boldsymbol{X})$ with

$$\|\boldsymbol{\beta}^0 - \boldsymbol{\beta}\| \le \frac{2\, N\, \delta}{c(\boldsymbol{X})}.$$

Assume the contrary: that for all $\boldsymbol{\beta} \in \widehat{\boldsymbol{\beta}}_1(\boldsymbol{y}, \boldsymbol{X})$ we have $\|\boldsymbol{\beta}^0 - \boldsymbol{\beta}\| > 2N\delta/c(\boldsymbol{X})$. Then $\boldsymbol{\beta}^0 \notin \widehat{\boldsymbol{\beta}}_1(\boldsymbol{y}, \boldsymbol{X})$. Since $\widehat{\boldsymbol{\beta}}_1(\boldsymbol{y}, \boldsymbol{X})$ is a convex and closed subset of $\mathbb{R}^p$, there exists $\boldsymbol{\beta} \in \widehat{\boldsymbol{\beta}}_1(\boldsymbol{y}, \boldsymbol{X})$ such that $\boldsymbol{\beta} + \alpha(\boldsymbol{\beta}^0 - \boldsymbol{\beta}) \notin \widehat{\boldsymbol{\beta}}_1(\boldsymbol{y}, \boldsymbol{X})$

for all $\alpha \in (0,1]$. For $\boldsymbol{b} = \boldsymbol{\beta}^0 - \boldsymbol{\beta}$ define $S_{\boldsymbol{b}}(\alpha)$ as above. Then we have $S_{\boldsymbol{b}}(\alpha) > S_{\boldsymbol{b}}(0)$ for all $\alpha \in (0,1]$ and therefore $S'_{\boldsymbol{b}}(0) > 0$. This implies

$$\begin{aligned}
\sum_{n=1}^{N} |y_n^0 - \boldsymbol{x}_n^T \boldsymbol{\beta}^0| &\geq \sum_{n=1}^{N} |y_n - \boldsymbol{x}_n^T \boldsymbol{\beta}^0| - \sum_{n=1}^{N} |y_n^0 - y_n| \\
&\geq S_{\boldsymbol{b}}(1) - N\delta \geq S_{\boldsymbol{b}}(0) + S'_{\boldsymbol{b}}(0) - N\delta \\
&\geq S_{\boldsymbol{b}}(0) + c(\boldsymbol{X})\,\|\boldsymbol{\beta}^0 - \boldsymbol{\beta}\| - N\delta \\
&> \sum_{n=1}^{N} |y_n - \boldsymbol{x}_n^T \boldsymbol{\beta}| + N\delta \\
&\geq \sum_{n=1}^{N} |y_n^0 - \boldsymbol{x}_n^T \boldsymbol{\beta}|
\end{aligned}$$

—a contradiction to $\boldsymbol{\beta}^0 \in \widehat{\boldsymbol{\beta}}_1(\boldsymbol{y}^0, \boldsymbol{X})$. □

Proof of Proposition 3. Set $M = \lfloor (1 + \min_{i=1,\dots,p} N_i)/2 \rfloor$. First assume that $r = 1$. Without loss of generality we can assume that $N_1 = \min_{1 \leq i \leq p} N_i$ and $\boldsymbol{x}_n = \boldsymbol{\chi}_1$ for $n = 1, \dots, N_1$. Then there exists $\beta \neq 0$ with $\boldsymbol{\chi}_i^T \boldsymbol{\beta} = 0$ for $i = 2, \dots, p$ such that

$$\sum_{n=M+1}^{N} |\boldsymbol{x}_n^T \boldsymbol{\beta}| - \sum_{n=1}^{M} |\boldsymbol{x}_n^T \boldsymbol{\beta}| = (N_1 - M)||\boldsymbol{\chi}_1^T \boldsymbol{\beta}| - M|\boldsymbol{\chi}_1^T \boldsymbol{\beta}| \leq 0.$$

Hence, we have $\mathcal{M}(\boldsymbol{X}, 1) \leq \lfloor (1 + \min_i N_i)/2 \rfloor$. Now consider any $\boldsymbol{\beta} \neq 0$ and any $E \subset \{1, \dots, N\}$ with $\operatorname{card} E = M - 1$. There exists $\alpha_1, \alpha_2, \dots, \alpha_p$ and $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_p \neq 0$ with $\boldsymbol{\chi}_i^T \boldsymbol{\beta}_j = 0$ for $i \neq j$ and $\boldsymbol{\beta} = \alpha_1 \boldsymbol{\beta}_1 + \dots + \alpha_p \boldsymbol{\beta}_p$ such that $|\boldsymbol{\chi}_i^T \boldsymbol{\beta}| = |\alpha_i|\,|\boldsymbol{\chi}_i^T \boldsymbol{\beta}_i|$ for $i = 1, \dots, p$. Let $j = \arg\max_{i=1,\dots,p} |\alpha_i|\,|\boldsymbol{\chi}_i^T \boldsymbol{\beta}_i|$. Then

$$\begin{aligned}
&\sum_{n \notin E} |\boldsymbol{x}_n^T \boldsymbol{\beta}| - \sum_{n \in E} |\boldsymbol{x}_n^T \boldsymbol{\beta}| \\
&\geq \sum_{i \neq j} N_i\,|\alpha_i|\,|\boldsymbol{\chi}_i^T \boldsymbol{\beta}_i| + (N_j - M + 1)|\alpha_j|\,|\boldsymbol{\chi}_j^T \boldsymbol{\beta}_j| \\
&\qquad - (M - 1)|\alpha_j|\,|\boldsymbol{\chi}_j^T \boldsymbol{\beta}_j| > 0.
\end{aligned}$$

Observing that $\mathcal{N}(\boldsymbol{X}) = N - N_1$ proves the proposition for $r = 1$. Proposition 1 provides the proposition for every $r \in [0,1]$. □

Proof of Proposition 5. *Equality.* For every E with $\operatorname{card} E = M \leq N/2$ the existence of $\boldsymbol{\beta}$ satisfying (14) is equivalent with the existence of $\boldsymbol{\beta}_1$ such that

$$\sum_{n \notin E} |\boldsymbol{\beta}_1 + t_n| \leq \sum_{n \in E} |\boldsymbol{\beta}_1 + t_n|. \tag{15}$$

But if $\boldsymbol{\beta}_1$ and E with card $E = M$ exist such that (15) is satisfied, then (15) holds either for $E = \{1, \ldots, M\}$ and $\boldsymbol{\beta}_0 = -t_N$ or for $E = \{N - M + 1, \ldots, N\}$ and $\boldsymbol{\beta}_0 = -t_1$.

Upper bound. Ellis and Morgenthaler showed in their Proposition 4.1 that $1/4$ is an upper bound for the breakdown point of the L_1 estimator if $N = 4K$ with $K \in \mathbb{N}$. Theorem 3 says that $\mathcal{M}(\boldsymbol{X}, 1)/N$ is the breakdown point of the L_1 estimator; hence, the proposed upper bound holds for $N = 4K$. To prove the assertion for $N = 4K + k$ with $k = 1, 2, 3$ assume that $\mathcal{M}(\boldsymbol{X}, 1) > \lfloor (N + 3)/4 \rfloor = K + 1$. Then the already proved first part of Proposition 5 (equality) yields for $M = K + 1$ that

$$\begin{aligned} \sum_{n=1}^{N-M} (t_n - t_1) &> \sum_{n=N-M+1}^{N} (t_n - t_1) \\ \text{and} \quad \sum_{n=M+1}^{N} (t_N - t_n) &> \sum_{n=1}^{M} (t_N - t_n). \end{aligned} \tag{16}$$

Let $L \in \{1, 2, 3\}$ such that $N + L = 4(K + 1)$. Consider the extended design $\boldsymbol{X}^* = (\boldsymbol{x}_1^*, \boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_{N+L}^*)^T$ given by $\boldsymbol{x}_n^* = (1, t_n^*)^T = (1, t_n)^T$ for $n = 1, \ldots, \lfloor (N + 1)/2 \rfloor$, $\boldsymbol{x}_n^* = (1, t_n^*)^T = (1, t_{\lfloor (N+1)/2 \rfloor})^T$ for $n = \lfloor (N + 1)/2 \rfloor + 1, \ldots, \lfloor (N + 1)/2 \rfloor + L$, and $\boldsymbol{x}_n^* = (1, t_n^*)^T = (1, t_{n-L})^T$ for $n = \lfloor (N + 1)/2 \rfloor + L + 1, \ldots, N + L$. This extended design also satisfies $t_1^* \leq t_2^* \leq \cdots \leq t_{N+L}^*$. Since $M = K + 1 \leq \lfloor (N + 1)/2 \rfloor \leq N - M$, we have by (16)

$$\begin{aligned} \sum_{n=1}^{N+L-M} (t_n^* - t_1^*) &= \sum_{n=1}^{N-M} (t_n - t_1) + \sum_{l=1}^{L} (t_{\lfloor (N+1)/2 \rfloor} - t_1) \\ &> \sum_{n=N-M+1}^{N} (t_n - t_1) = \sum_{n=N+L-M+1}^{N+L} (t_n^* - t_1^*) \end{aligned}$$

and

$$\begin{aligned} \sum_{n=M+1}^{N+L} (t_{N+L}^* - t_n^*) &= \sum_{n=M+1}^{N} (t_N - t_n) + \sum_{l=1}^{L} (t_N - t_{\lfloor (N+1)/2 \rfloor}) \\ &> \sum_{n=1}^{M} (t_N - t_n) = \sum_{n=1}^{M} (t_{N+L}^* - t_n^*). \end{aligned}$$

This implies, in view of the equality part of Proposition 5, that $\mathcal{M}(\boldsymbol{X}^*, 1) > K + 1$. But since $N + L = 4(K + 1)$, we obtain from the result of Ellis and Morgenthaler (1992) the contradiction: $\mathcal{M}(\boldsymbol{X}^*, 1) \leq K + 1$. □

Proof of Proposition 6. Demanding the existence of $\boldsymbol{\beta} = \boldsymbol{A}^T \tilde{\boldsymbol{\beta}}$ for some $\tilde{\boldsymbol{\beta}} \in \mathbb{R}^2$ such that (14) holds for $\boldsymbol{\beta}$ is a more stringent requirement than demanding an existence of a general $\boldsymbol{\beta}$ satisfying (14). □

Appendix: computation of $\mathcal{M}(\boldsymbol{X}, 1)$

For computing $\mathcal{M}(\boldsymbol{X}, 1)$ numerically, the following observations are useful. For $E \subseteq \{1, 2, \ldots, N\}$, let

$$F_E(\boldsymbol{\beta}) = \frac{\sum_{n \in E} |\boldsymbol{x}_n^T \boldsymbol{\beta}|}{\sum_{n=1}^N |\boldsymbol{x}_n^T \boldsymbol{\beta}|}.$$

Note that if $\mathcal{N}(\boldsymbol{X}) < N$, then $\mathcal{M}(\boldsymbol{X}, 1) - 1$ is equal to the maximal cardinality of a set $E \subset \{1, 2, \ldots, N\}$ such that $\max_{\boldsymbol{\beta}} F_E(\boldsymbol{\beta}) < 1/2$. For $c \neq 0$, $F_E(c\boldsymbol{\beta}) = F_E(\boldsymbol{\beta})$; thus, since we can restrict the domain of maximization to a compact set

$$A = \left\{ \boldsymbol{\beta} : \sum_{n=1}^N |\boldsymbol{x}_n^T \boldsymbol{\beta}| = 1 \right\},$$

the maximum is attained for every E. Let $\mathcal{X}$ be *the set* of all design points $\boldsymbol{x}_n$ (that is, multiple design points enter $\mathcal{X}$ only once). Let $\mathcal{D}$ be the set of all subsets D of $\mathcal{X}$ with the property that the linear space spanned by the elements of D is $(p-1)$-dimensional. Let

$$D^\perp = \{\boldsymbol{\beta}\colon \boldsymbol{x}_n^T \boldsymbol{\beta} = 0 \text{ for all } \boldsymbol{x}_n^T \in D\}$$

be the orthogonal complement of this space and let

$$\mathcal{B} = \bigcup_{D \in \mathcal{D}} (A \cap D^\perp).$$

Now, let $\mathcal{S} = \{-1, 1\}^N$. For $\boldsymbol{s} \in \mathcal{S}$ we define

$$A_{\boldsymbol{s}} = \left\{ \boldsymbol{\beta} \in A\colon |\boldsymbol{x}_n^T \boldsymbol{\beta}| = s_n \boldsymbol{x}_n^T \boldsymbol{\beta} \text{ for } n = 1, 2, \ldots, N \right\}.$$

In other terms, if $\boldsymbol{\beta} \in A_{\boldsymbol{s}}$, then $\operatorname{sign}(\boldsymbol{x}_n^T \boldsymbol{\beta}) = s_n$ or 0.

Lemma 7. *For every $\boldsymbol{s} \in \mathcal{S}$, the set $A_{\boldsymbol{s}}$ is compact and convex (possibly empty). All extremal points of $A_{\boldsymbol{s}}$ lie in $\mathcal{B}$; F_E is linear on $A_{\boldsymbol{s}}$ for every E.*

Proof. Since $A_{\boldsymbol{s}} \subset A$, it is bounded; the properties determining $A_{\boldsymbol{s}}$ preserve under a passage to limit, hence $A_{\boldsymbol{s}}$ is closed. Convex combinations also remain in $A_{\boldsymbol{s}}$; since

$$\sum_{n=1}^N |\boldsymbol{x}_n^T \boldsymbol{\beta}| = \sum_{n=1}^N s_n \boldsymbol{x}_n^T \boldsymbol{\beta},$$

all convex combinations of elements of $A_{\boldsymbol{s}}$ belong to A. Linearity of F_E on $A_{\boldsymbol{s}}$ follows in an analogous way.

Fix $\boldsymbol{s} \in \mathcal{S}$. Let $\boldsymbol{\beta}$ be an extremal point of $A_{\boldsymbol{s}}$ and let C be the set of all those $\boldsymbol{x}_n$ for which $\boldsymbol{x}_n^T\boldsymbol{\beta} = 0$. Suppose that the dimension of the linear space spanned by C does not exceed $p-2$. Then the linear space spanned by $C \cup \{\sum_{n=1}^N s_n\boldsymbol{x}_n\}$ has dimension at most $p-1$; consequently, there is a $\tilde{\boldsymbol{\beta}} \neq \mathbf{0}$ such that $\boldsymbol{x}_n^T\tilde{\boldsymbol{\beta}} = 0$ for all $\boldsymbol{x}_n \in C$ and $\sum_{n=1}^N s_n\boldsymbol{x}_n^T\tilde{\boldsymbol{\beta}} = 0$. For all $\boldsymbol{x}_n \in \mathcal{X} \setminus C$, $|\boldsymbol{x}_n^T\boldsymbol{\beta}| \geq \eta$ for some $\eta > 0$; we can suppose that $|\boldsymbol{x}_n^T\tilde{\boldsymbol{\beta}}| \leq \eta/2$ for all $n = 1, 2, \ldots, N$, taking a suitable nonzero multiple of $\tilde{\boldsymbol{\beta}}$ otherwise. As a consequence,

$$\operatorname{sign}(\boldsymbol{x}_n^T\boldsymbol{\beta} + \boldsymbol{x}_n^T\tilde{\boldsymbol{\beta}}) = \operatorname{sign}(\boldsymbol{x}_n^T\boldsymbol{\beta}) = \operatorname{sign}(\boldsymbol{x}_n^T\boldsymbol{\beta} - \boldsymbol{x}_n^T\tilde{\boldsymbol{\beta}}) \tag{17}$$

for all $\boldsymbol{x}_n \in \mathcal{X} \setminus C$, and also for all $\boldsymbol{x}_n \in C$. Therefore,

$$\begin{aligned} \sum_{n=1}^N \left|\boldsymbol{x}_n^T(\boldsymbol{\beta} + \tilde{\boldsymbol{\beta}})\right| &= \sum_{n=1}^N s_n\boldsymbol{x}_n^T(\boldsymbol{\beta} + \tilde{\boldsymbol{\beta}}) \\ &= \sum_{n=1}^N s_n\boldsymbol{x}_n^T\boldsymbol{\beta} + \sum_{n=1}^N s_n\boldsymbol{x}_n^T\tilde{\boldsymbol{\beta}} = 1, \end{aligned}$$

and analogously $\sum_{n=1}^N |\boldsymbol{x}_n^T(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})| = 1$. We obtain, in view of (17), that β is a convex combination of $\boldsymbol{\beta} + \tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}$, both elements of $A_{\boldsymbol{s}}$. This contradicts the extremality of $\boldsymbol{\beta}$; hence, the space spanned by C has dimension at least $p-1$. It cannot have dimension p since then $\boldsymbol{x}_n^T\boldsymbol{\beta} = 0$ for all $\boldsymbol{x}_n \in C$ would imply $\boldsymbol{\beta} = \mathbf{0}$, a contradiction with the fact that $\boldsymbol{\beta} \in A$. Thus, the dimension is $p-1$, as had to be proved. □

Proposition 7. *For every $E \subseteq \{1, 2, \ldots, N\}$, the maximum $\max_{\boldsymbol{\beta}} F_E(\boldsymbol{\beta})$ is attained on some point of $\mathcal{B}$, which satisfies*

$$\frac{1}{2}\operatorname{card}\mathcal{B} \leq \binom{\operatorname{card}\mathcal{X}}{p-1}. \tag{18}$$

Proof. Since $D^\perp$ is a one-dimensional subspace of $\mathbb{R}^p$, the set $A \cap D^\perp$ contains, for any $D \in \mathcal{D}$, at most two elements; this proves (18). The remaining part of the statement follows from Lemma 7 and the fact that $A = \bigcup_{\boldsymbol{s}\in\mathcal{S}} A_{\boldsymbol{s}}$—since the maximum of a linear function over a compact convex set is attained in one of its extremal points. □

Proposition 7 says that for computing $\mathcal{M}(\boldsymbol{X}, 1)$, we have only to inspect $\boldsymbol{\beta} \in \mathcal{B}$. Note that, in fact, we have to consider only one $\boldsymbol{\beta}$ from each $A \cap D^\perp$, the second one is $-\boldsymbol{\beta}$, with the same $|\boldsymbol{x}_n^T\boldsymbol{\beta}|$'s for all $n = 1, 2, \ldots, N$.

Acknowledgements

We are very grateful to Jana Jurečková for valuable suggestions and discussions, and to her, Xuming He and Stephen Portnoy for exhaustive replies to our queries. The authors express their thanks to Viktor Kurotschka and Andrej Pázman and to the Deutsche Forschungsgemeinschaft and Slovak Grant Agency for Science for supporting their mutual visits of the Free University and Comenius University.

References

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular Variation.* Cambridge University Press, Cambridge.

Ellis, S. P. and Morgenthaler, S. (1992). Leverage and breakdown in L_1 regression. *J. Amer. Statist. Assoc.* **87**, 143–148.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics—The Approach Based on Influence Functions.* Wiley, New York.

He, X., Jurečková, J., Koenker, R. and Portnoy, S. (1990). Tail behavior of regression estimators and their breakdown points. *Econometrica* 58, 1195–1214.

Hoaglin, D. C., Mosteller, F. and Tukey, J. W. (eds.) (1983). *Understanding Robust and Exploratory Data Analysis.* Wiley, New York.

Huber, P. J. (1981). *Robust Statistics.* Wiley, New York.

Huber, P. J. (1984). Finite sample breakdown of M- and P-estimators. *Ann. Statist.* **12**, 119–126.

Jurečková, J. (1981). Tail-behavior of location estimators. *Ann. Statist.* **9**, 578–585.

Maronna, R. A., Bustos, O. H. and Yohai, V. J. (1979). Bias- and efficiency-robustness of general M-estimators for regression with random carriers. In: *Smoothing Techniques for Curve Estimation,* eds. T. Gasser and M. Rosenblatt, Lecture Notes in Mathematics **757**, Springer, Berlin, 91–116.

Maronna, R. A., Yohai, V. J. and Zamar, R. J. (1993). Bias-robust regression estimation: A partial survey. In: *New Directions in Statistical Data Analysis and Robustness,* eds. S. Morgenthaler, E. Ronchetti and W. A. Stahel, Birkhäuser, Basel, 157–176.

Mili, L. and Coakley, C. W. (1993). Robust estimation in structured linear regression. *Technical Report No. 93-13,* Department of Statistics, Virginia Polytechnic Institute and State University. (To appear in Ann. Statist.)

Mizera, I. (1996). Weak continuity of redescending M-estimators of location with an unbounded objective function. In: *PROBASTAT'94, Proceedings of the Second International Conference on Mathematical Statistics,* Tatra Mountains Mathematical Publications **7**, 343–347.

Morgenthaler, S. and Tukey, J. W. (1991). *Configural Polysampling.* Wiley, New York.

Müller, Ch. H. (1995). Breakdown points for designed experiments. *J. Statist. Plann. Inference.* **45**, 413–427.

Müller, Ch. H. (1996a). High breakdown point designs. In: *Robust Statistics, Data Analysis, and Computer Intensive Methods, In Honor of Peter Huber's 60th Birthday,* ed. H. Rieder, Lecture Notes in Statistics **109**, Springer, New York, 353–360.

Müller, Ch. H. (1996b). Optimal breakdown point maximizing designs. In: *PROBASTAT'94, Proceedings of the Second International Conference on Mathematical Statistics,* Tatra Mountains Mathematical Publications **7**, 79–85.

Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes.* Springer, New York.

Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* Wiley, New York.

Comenius University
Department of Probability
and Statistics, Mlynská dolina
SK-84215 Bratislava
Slovakia
e-mail: mizera@fmph.uniba.sk

Free University of Berlin
Department of Mathematics
Arnimallee 2–6
D-14195 Berlin
Germany
e-mail: mueller@math.fu-berlin.de