Journal de la Société Française de Statistique
Vol. 153 No. 1 (2012)

Tail-behavior of estimators and of their one-step versions. *

Jana Jurečková [1]

Abstract: The finite-sample breakdown points and finite-sample tail behavior are studied for a class of equivariant estimators in the linear regression model under a fixed design. The same is considered for the one-step and k-step versions of the estimators, starting with an initial estimator. It is shown that the tail-behavior of the one- and k-step versions of an estimator is determined mainly by that of the initial estimator.

Résumé : Les points de rupture et le comportement des queues sous des échantillons finis sont etudiés pour une classe d'estimateurs equivariants dans le modèle linéaire avec un design fixe. Des résultats du même type sont obtenus pour des itérations de type Newton-Raphson d'un estimateur initial. On démontre que le comportement des queues de ces estimateurs itérés est principalemnt déterminé par celui de l'estimateur initial.

Keywords: breakdown point, equivariant estimator, tail-behavior
Mots-clés : point de rupture, estimateur equivariant, comportement des queues
AMS 2000 subject classifications: 35L05, 35L70

1. Introduction

Consider the linear regression model

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{U} \tag{1.1}$$

with vector of observations $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$, matrix $\mathbf{X} = \left\|\mathbf{x}_1, \ldots, \mathbf{x}_n\right\|$ of order $n \times p$ with the full rank p, parameter $\beta = (\beta_1, ..., \beta_p)^\top$ and with the vector $\mathbf{U} = (U_1, ..., U_n)^\top$ of i.i.d. errors, distributed according to the distribution function F with density f, which is absolutely continuous and has a non-vanishing characteristic function φ.

The problem is estimating β under the loss function $L(\mathbf{b}, \beta) = L(\mathbf{b} - \beta)$. The corresponding risk $R(\mathbf{T}, \beta)$ is invariant with respect to the group of transformations $\mathscr{G} = \{\mathbf{Y} + \mathbf{X}b, \ \mathbf{b} \in \mathbb{R}^p\}$, and one possible maximal invariant for $\mathscr{G}$ is the vector $\mathbf{Z} = \mathbf{Y} - \widehat{\mathbf{Y}}$ with $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\beta}$ and $\widehat{\beta} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$ being the LSE of β. Under the invariant risk, it is natural to restrict considerations to equivariant estimators $\mathbf{T}_n$ satisfying $\mathbf{T}_n(\mathbf{Y} + \mathbf{X}\mathbf{b}) = \mathbf{T}_n(\mathbf{Y}) + \mathbf{b}$. Every equivariant estimator $\mathbf{T}_n$ we obtain from an initial equivariant estimator $\mathbf{T}_n^0$ with a finite risk, adding an invariant statistic, i.e. adding any possible (vector) function of $\mathbf{Z}$:

$$\mathbf{T}_n = \mathbf{T}_n^0 + \mathbf{v}(\mathbf{Z}).$$

* The research was supported by the Grant GAC(R 201/12/0083.
[1] Department of Probability and Statistics, Charles University in Prague, Sokolovská 83, CZ-186 75 Prague 8, Czech Republic.
E-mail: jurecko@karlin.mff.cuni.cz

For instance, the minimum risk equivariant estimator with respect to the quadratic loss $\|\mathbf{b}-\beta\|^2$ is equal to $\mathbf{T}_n^* = \mathbf{T}_n^0 - E_{\mathbf{0}}(\mathbf{T}_n^0|\mathbf{Z})$. To calculate it, we must know F, but even then its calculation is rather technical. Possible approximations of $\mathbf{T}_n^*$ were studied in [4].

Many robust estimators in the linear model, as M-, L- and R-estimators, correspond to an invariant loss and so are regression equivariant. Also their calculation is often technical, and they are often approximated by their one-step versions. These are the first Newton-Raphson iterations of the estimating equation, starting with an initial estimator, even if the estimator itself is not exactly the root. The k-step versions are also considered, and many results in the literature show the asymptotic closeness of $\mathbf{T}_n$ and of its one-step version.

However, though the estimators are asymptotically normal, some of their properties are typically finite-sample, despite of the widely accepted view that many properties are inherited by the asymptotic normality. The admissibility of an estimator with respect a specified loss function is a typical example. Generally accepted view was that, while the classical procedures are sensitive to the heavy-tailed distributions, this problem is solved by using the robust procedures. This is true only to some extent. It was shown in [5] that the distribution of an equivariant estimator of location is heavy-tailed for any finite n provided the parent distribution is heavy tailed. Although the Pareto index is increasing with n, it is always finite and the distribution never gets exponentially tailed.

We shall illustrate this problem, using the tail-behavior measure for regression estimators proposed in [1]. Under some conditions, we derive the lower bound for the tail behavior of M-estimators in model (1.1). Then we shall consider the one-step versions in model (1.1) and show that the tail-behavior of the Newton-Raphson iteration is determined mainly by that of the initial estimator, and this holds for the k-step version for any finite $k \geq 1$.

2. Tail behavior of equivariant estimators

A possible measure of tail performance of estimator T_n in the location model with $Y_1, \ldots, Y_n$ independent and identically distributed according a symmetric $F(y-\theta)$ is

$$B(a, T_n) = \frac{-\log\ P_\theta(T_n - \theta > a)}{-\log\ (1-F(a))} \quad \text{for fixed } n \quad \text{as } a \to \infty \tag{2.1}$$

(see [2]). Then $1 \leq \liminf_{a\to\infty} B(a,T_n) \leq \limsup_{a\to\infty} B(a,T_n) \leq n$ for any equivariant T_n satisfying weak conditions, and both bounds are attainable by the sample mean $T_n = \bar{X}_n$; the upper one for the exponentially tailed, the lower one for heavy tailed distributions.

Remind a close link between the tail performance and the breakdown point for a large class of location estimators, proved in [1]:

Theorem 2.1. *(He et al. (1990)). Let $T(Y_1, \ldots, Y_n)$ be a location equivariant estimator of θ, nondecreasing in each argument Y_i. Then T_n has a universal breakdown point m^* and*

$$m^* \leq \liminf_{a\to\infty} B(a,T_n) \leq \limsup_{a\to\infty} B(a,T_n) \leq n - m^* + 1 \tag{2.2}$$

is valid for any symmetric, absolutely continuous F of Y_1 satisfying

$$\lim_{z\to\infty} \frac{\log(1-F(z+c))}{\log(1-F(z))} = 1 \quad \textit{for any fixed } c > 0. \tag{2.3}$$

Theorem 2.1 holds both for exponential and algebraic tails of F. For example, the median T_n has breakdown $m^* = \frac{n}{2}$ and for n odd $P_\theta(T_n - \theta > a)$ tends to zero as $a \to \infty$ just $\frac{n+1}{2}$-times faster than the tails of the underlying error distribution.

The measure of tail performance (2.1) was extended in [1] to the linear model (1.1) in the following way:

$$B(a, \mathbf{T}_n) = \frac{-\ln P_\beta \left\{ \max_{1\le i\le n} |\mathbf{x}_i^\top (\mathbf{T}_n - \beta)| > a \right\}}{-\ln(1 - F(a))} \quad \text{for } a \gg 0. \tag{2.4}$$

He at al. (1990) showed that $\limsup_{a\to\infty} B(\mathbf{T}_n, a) \le n$ for a regression equivariant estimator under a weak condition, and showed that $\limsup_{a\to\infty} B(\widehat{\beta}_n, a) \le n/p$ for the LSE and the normal distribution. However, this upper bound holds only under a balanced design, while generally we have

$$\lim_{a\to\infty} B(\widehat{\beta}_n, a) = 1/\hat{h}_n \quad \text{for the normal } F$$
$$\lim_{a\to\infty} B(\widehat{\beta}_n, a) = 1 \quad \text{for the heavy-tailed } F$$

where $\hat{h}_n$ is the maximal diagonal element of the hat matrix $\mathbf{H} = \mathbf{X}_n(\mathbf{X}_n^\top \mathbf{X}_n)^{-1}\mathbf{X}_n^\top$. It indicates that the matrix $\mathbf{X}_n$ seriously affects the tail behavior of an estimator in model (1.1). A bad message for an estimator is the low value of $B(\mathbf{T}_n, a)$, which hints that the estimator utilizes only a small proportion of the data. He at al. (1990) derived the lower bound of the tail behavior for the L_1-estimator and for some M-estimators of β with the criterion function $\rho(x)$ close to $|x|$.

We shall derive the lower bounds of the tail behavior for a more general class of M-estimators, including the Huber-type estimators and some redescending M-estimators, defined as

$$\mathbf{T}_n = \arg\min_{\mathbf{b}\in\mathbb{R}^p} \left\{ \sum_{i=1}^n \rho(Y_i - \mathbf{x}_i^\top \mathbf{b}) \right\} \tag{2.5}$$

We shall show that the tail behavior of such estimators has a lower bound > 1 for both heavy-tailed and light-tailed F, satisfying

$$\begin{aligned} &0 < F(z) < 1, \quad F(z) + F(-z) = 1, \quad z \in \mathbb{R}^1 \\ \text{and} \quad & \\ &\lim_{a\to\infty} \frac{-\ln(1 - F(a + c))}{-\ln(1 - F(a))} = 1 \quad \text{for } \forall c > 0. \end{aligned} \tag{2.6}$$

Following Mizera and Müller (1999), we shall impose the following conditions on the criterion function ρ, discussed in [8] in detail:

(i) ρ is absolutely continuous, nondecreasing on $[0, \infty)$ and $\rho(z) \ge 0$, $\rho(z) = \rho(-z)$, $z \in \mathbb{R}^1$.

(ii) $\rho(z)$ is unbounded and its derivative $\psi(z)$ is bounded for $z \in \mathbb{R}^1$.

(iii) ρ is subadditive in the sense that there exists $L > 0$ such that $\rho(z_1 + z_2) \le \rho(z_1) + \rho(z_2) + L$ for $z_1, z_2 \ge 0$.

Given ρ satisfying (i)–(iii), define

$$m_* = m_*(n, \mathbf{X}, \rho) = \min\left\{\text{card}\, M : \sum_{i\in M} \rho(\mathbf{x}_i^\top \mathbf{b}) \geq \sum_{i\notin M} \rho(\mathbf{x}_i^\top \mathbf{b}) \quad \text{for some } \mathbf{b} \neq \mathbf{0}\right\} \tag{2.7}$$

where M runs over the subsets of $N = \{1, 2, \ldots, n\}$. The following theorem shows that m_* is the lower bound for the tail behavior of M-estimator generated by ρ :

Theorem 2.2. *Under F satisfying (2.6) and ρ satisfying (i)–(iii), the tail behavior of M-estimator defined in (2.5) satisfies*

$$\liminf_{a\to\infty} B(\mathbf{T}_n, a) \geq m_* \tag{2.8}$$

with m_* defined in (2.7).

Proof. Regarding the equivariance of the M-estimators in (2.5), we can assume $\beta = \mathbf{0}$ without loss of generality. L will be a generic constant. By (i)–(iii),

$$\begin{aligned}\sum_N \rho(y_i - \mathbf{x}_i^\top \mathbf{b}) &\geq \sum_{N\setminus M} \rho(\mathbf{x}_i^\top \mathbf{b}) - \sum_{N\setminus M} \rho(y_i) - \left(\sum_M \rho(\mathbf{x}_i^\top \mathbf{b}) - \sum_M \rho(y_i)\right) - L \\ &= \sum_{N\setminus M} \rho(\mathbf{x}_i^\top \mathbf{b}) - \sum_M \rho(\mathbf{x}_i^\top \mathbf{b}) + \sum_N \rho(y_i) - 2\sum_{N\setminus M} \rho(y_i) - L.\end{aligned}$$

By the definition of m_*, there exists $\varepsilon > 0$ such that

$$\sum_M \rho(\mathbf{x}_i^\top \mathbf{b}) \leq (1-\varepsilon) \sum_{N\setminus M} \rho(\mathbf{x}_i^\top \mathbf{b}) \tag{2.9}$$

for every $M \subset N$ of size $m = m_* - 1$ and every $\mathbf{b} \in \mathbb{R}^p$. Assume that all but $m = m_* - 1$ of the $\rho(y_i)$'s are uniformly bounded. Because (2.9) implies

$$\sum_N \rho(y_i - \mathbf{x}_i^\top \mathbf{b}) - \sum_N \rho(y_i) \geq \varepsilon \sum_{N\setminus M} \rho(\mathbf{x}_i^\top \mathbf{b}) - 2\sum_{N\setminus M} \rho(y_i) - L,$$

we deduce that there exists $C > 0$ such that if $\|\mathbf{b}\| > C$ then $\sum_N \rho(y_i - \mathbf{x}_i^\top \mathbf{b}) - \sum_N \rho(y_i) > 0$ and $\mathbf{b}$ cannot be a minimizer of $\sum_N \rho(y_i - \mathbf{x}_i^\top \mathbf{t})$. Hence, for $\left[\max_i \rho(\mathbf{x}_i^\top \mathbf{T}_n) \gg 1\right]$ we need to have the $\rho(y_i)$ unbounded for $i \in N \setminus M$ and $\max_i \rho(\mathbf{x}_i^\top \mathbf{T}_n) \leq K\rho(|y|_{n:n-m_*+1})$, where $|y|_{n:n-m_*+1}$ is the $(n - m_* + 1)$-th order statistic of $|y_1|, \ldots, |y_n|$. Thus,

$$P\left\{\max_i |\mathbf{x}_i^\top \mathbf{T}_n| > a\right\} \leq P\left\{|y|_{n:n-m_*+1} \geq K'a\right\}.$$

Using the distribution of order statistics of $|Y_1|, \ldots, |Y_n|$ for F satisfying (2.6), we conclude that $\liminf_{a\to\infty} B(\mathbf{T}_n) \geq m_*$. □

Remark 2.1. *The lower bound in (2.8) coincides with the lower bound derived by Mizera and Müller (1999) for the finite-sample breakdown point of the M-estimator $\mathbf{T}_n$. Moreover, if ρ is regularly varying at infinity with an exponent $r \geq 0$, the lower bound in [8] is further modified to a form depending on r.*

3. One-step version of $\mathbf{T}_n$

A broad class of estimators $\mathbf{T}_n$ of β admit a representation

$$\mathbf{T}_n(\mathbf{Y}) = \beta + \frac{1}{\gamma}(\mathbf{X}_n^\top \mathbf{X}_n)^{-1} \sum_{i=1}^n \mathbf{x}_i \psi(Y_i - \mathbf{x}_i^\top \beta) + \mathbf{R}_n, \; \|\mathbf{R}_n\| = o_p(\|\mathbf{X}_n^\top \mathbf{X}_n\|^{-1/2}) \tag{3.1}$$

with a suitable function ψ and a functional $\gamma = \gamma(\psi, F)$. The representation (3.1) is valid e.g. for an M-estimator (2.5). The representation (3.1) holds as an identity if $\mathbf{T}_n$ is the least squares estimator $\widehat{\beta}_n$ and ψ is linear. By Kagan et al. (1973), the admissibility of $\widehat{\beta}_n$ with respect to the quadratic loss implies that $E_{\mathbf{0}}(\widehat{\beta}_n|\mathbf{Z}_n) = \mathbf{0}$ and this in turn implies the normality of the distribution.

The one-step version of $\mathbf{T}_n$ with representation (3.1) is defined as the one-step Newton-Raphson iteration of the system of equations

$$\sum_{i=1}^n \mathbf{x}_i \psi(Y_i - \mathbf{x}_i^\top \mathbf{b}) = \mathbf{0} \tag{3.2}$$

even when the global minimum of (2.5) is not a root of (3.2), as in the case of L_1-estimator or of other M-estimators with discontinuous ψ.

For simplicity, standardize the sequence $\{\mathbf{X}_n\}$ so that

$$\lim_{n\to\infty} \mathbf{Q}_n^* = \mathbf{Q}^*, \quad \mathbf{Q}_n^* = n^{-1}\mathbf{X}_n^\top \mathbf{X}_n, \tag{3.3}$$

and assume that $\mathbf{Q}^*$ is a positively definite $p \times p$ matrix.

Start with an initial estimator $\mathbf{T}_n^{(0)}$ of β satisfying $n^{1/2}(\mathbf{T}_n^{(0)} - \beta) = O_p(1)$. Assume that $\gamma \neq 0$ and let $\hat{\gamma}_n$ be a consistent estimator of γ such that $1 - (\gamma/\hat{\gamma}_n) = O_p(n^{-1/2})$. Then the one-step version of $\mathbf{T}_n$ is defined as

$$\mathbf{T}_n^{(1)} = \begin{cases} \mathbf{T}_n^{(0)} + (n\hat{\gamma}_n)^{-1}(\mathbf{Q}_n^*)^{-1} \sum_{i=1}^n \mathbf{x}_i \psi(Y_i - \mathbf{x}_i^\top \mathbf{T}_n^{(0)}) & \ldots \quad \text{if } \hat{\gamma}_n \neq 0 \\ \mathbf{T}_n^{(0)} & \ldots \quad \text{otherwise} \end{cases} \tag{3.4}$$

The two-step or the k-step versions of $\mathbf{T}_n$ are defined analogously for $k = 2, 3, \ldots$. For a studentized M-estimator $\mathbf{M}_n$ of β defined as

$$\mathbf{T}_n = \arg\min \left\{ \sum_{i=1}^n \rho\left(\frac{Y_i - \mathbf{X}_i^\top \mathbf{b}}{S_n}\right), \, \mathbf{b} \in \mathbb{R}^p \right\},$$

its one-step version is the one-step Newton-Raphson iteration of the system of equations

$$\sum_{i=1}^n \mathbf{x}_i \psi\left(\frac{Y_i - \mathbf{x}_i^\top \mathbf{b}}{S_n}\right) = \mathbf{0}, \quad \mathbf{b} \in \mathbb{R}^p \tag{3.5}$$

where $S_n(\mathbf{Y}_n)$ is a studentizing scale statistic such that

$$S_n(\mathbf{y}) > 0, \; S_n(c(\mathbf{y} + \mathbf{X}\mathbf{b})) = cS_n(\mathbf{y}), \; c > 0, \; \mathbf{y} \in \mathbb{R}^n, \; \mathbf{b} \in \mathbf{R}^p,$$

and to which exists a functional $S = S(F) > 0$ such that $n^{1/2}(S_n - S(F)) = O_p(1)$ as $n \to \infty$. Under some conditions on ψ and on F, the one-step version of $\mathbf{T}_n$ was modified in [6] so that $\mathbf{T}_n^{(1)}$ inherited the breakdown point of the initial estimator $\mathbf{T}_n^{(0)}$ and the asymptotic efficiency of $\mathbf{T}_n$. Namely, under F symmetric and $\psi = \rho'$ skew-symmetric, absolutely continuous, non-decreasing and bounded, and under some other smoothness conditions, we define the one-step version in the following way:

$$\mathbf{T}_n^{(1)} = \begin{cases} \mathbf{T}_n^{(0)} + \hat{\gamma}_n^{-1}\mathbf{W}_n & \ldots \quad \|\hat{\gamma}_n^{-1}\mathbf{W}_n\| \le c,\ 0 < c < \infty \\ \mathbf{T}_n^{(0)} & \ldots \quad \text{otherwise} \end{cases} \tag{3.6}$$

where $\mathbf{W}_n = n^{-1}\mathbf{Q}_n^{*-1}\sum_{i=1}^n \mathbf{x}_i \psi\left(\frac{Y_i - \mathbf{x}_i^\top \mathbf{T}_n^{(0)}}{S_n}\right)$ and

$$\hat{\gamma}_n = \frac{1}{2\sqrt{n}}\sum_{i=1}^n x_{i1}\left[\frac{\psi(Y_i - \mathbf{x}_i^\top \mathbf{T}_n^{(0)} + n^{-1/2}\mathbf{x}_i^\top \mathbf{q}_n^{(1)})}{S_n} - \frac{\psi(Y_i - \mathbf{x}_i^\top \mathbf{T}_n^{(0)} - n^{-1/2}\mathbf{x}_i^\top \mathbf{q}_n^{(1)})}{S_n}\right] \tag{3.7}$$

where $\mathbf{q}_n^{(1)}$ is the first column of $\mathbf{Q}_n^{*-1}$. Then

Theorem 3.1. *(Jurečková and Portnoy 1987). Under the above conditions,*

(i) $$\|\mathbf{T}_n^{(1)} - \mathbf{T}_n\| = O_p(n^{-1}).$$

(ii) If $\mathbf{T}_n^{(0)}$ *has finite sample breakdown point* m_n*, then* $\mathbf{T}_n^{(1)}$ *has the same breakdown point* m_n*.*

(iii) If $\mathbf{T}_n^{(0)}$ *is affine equivariant, then* $P\left\{\mathbf{T}_n^{(1)}(\mathbf{X}_n\mathbf{A}) \neq \mathbf{A}^{-1}\mathbf{T}_n^{(1)}(\mathbf{X}_n)\right\} \to 0$ *as* $n \to \infty$ *for any regular* $p \times p$ *matrix* $\mathbf{A}$*.*

Remark 3.1. *The results are true even for the initial estimator satisfying* $\|\mathbf{T}_n^{(0)} - \beta\| = O_p(n^{-\tau})$ *for* τ *satisfying* $\frac{1}{4} < \tau \le \frac{1}{2}$*.*

Remark 3.2. *Starting with an estimator with a low rate of consistency needs many observations to achieve a desired precision. In a model with a scalar parameter, it was proved in [7] that* $|T_n - T_n^{(1)}| = o_p(n^{-1})$ *only in a symmetric location model (*F *symmetric and* ψ *skew-symmetric) or generally only if the initial* $T_n^{(0)}$ *has the same influence function as* T_n*. The rate of approximation of the k-step version* $T_n^{(k)}$*,* $k \ge 2$*, to* T_n *depends on the smoothness of* ψ *: while* $|T_n - T_n^{(k)}| = O_p(n^{-\frac{1}{2}-\frac{k}{2}})$ *for an absolutely continuous* ψ*, it is only* $|T_n - T_n^{(k)}| = O_p\left(n^{-1+2^{-k-1}}\right)$ *for* ψ *with jump-discontinuities (see [3]).*

We shall show that in the location model not only the breakdown point but also the tail-behavior of the one-step version is determined by that of the initial estimate. It all leads to a conjecture that the finite-sample properties of $T_n^{(1)}$ depend on the properties of $T_n^{(0)}$, while the asymptotic properties depend on those of the non-iterated estimator T_n.

3.1. Tail behavior of one-step and k-step versions in the location model

Let us now consider the equivariant estimator T_n of location parameter, satisfying the representation (3.1), and its modified one-step version with an equivariant initial estimator $T_n^{(0)}$:

$$T_n^{(1)} = \begin{cases} T_n^{(0)} + \hat{\gamma}_n^{-1} W_n & \dots \text{ if } |\hat{\gamma}_n^{-1} W_n| \leq c, \ 0 < c < \infty \\ T_n^{(0)} & \dots \text{ otherwise} \end{cases} \tag{3.8}$$

where $W_n = n^{-1} \sum_{i=1}^n \psi(Y_i - T_n^{(0)}) = O_p(n^{-1/2})$ and with $\hat{\gamma}_n$ of type (3.7). Define the k-step version of T_n analogously. Then $P_\theta(T_n \neq T_n^{(0)} + \hat{\gamma}_n^{-1} W_n) \to 0$ as $n \to \infty$, hence $T_n^{(1)} - T_n = o_p(n^{-1/2})$. If $T_n^{(0)}$ is equivariant, so is $T_n^{(1)}$, because both $\hat{\gamma}_n$ and W_n are invariant. Surprisingly, the tail behavior of $T_n^{(1)}$ and of $T_n^{(k)}$ depends more on that of $T_n^{(0)}$ than on the tail-behavior of non-iterative T_n. This situation is described in the following theorem:

Theorem 3.2. *Let $Y_1, \dots, Y_n$ be a sample from a population with d.f. $F(y - \theta)$, F symmetric and satisfying (2.3). Let T_n be an equivariant estimator of θ admitting the representation (3.1) with a bounded skew-symmetric non-decreasing ψ. Then*

$$\liminf_{a \to \infty} B(T_n^{(0)}, a) \leq \liminf_{a \to \infty} B(T_n^{(k)}, a) \leq \limsup_{a \to \infty} B(T_n^{(k)}, a) \leq \limsup_{a \to \infty} B(T_n^{(0)}, a)$$

for $k = 1, 2, \dots$.

Proof: Let us prove the proposition for $k = 1$; the case $k > 1$ is analogous. The tail behavior measure of $T_n^{(1)}$ satisfies, for $0 < \delta < 1$ and sufficiently large a,

$$\begin{aligned} P_0(T_n^{(1)} > a) &\leq P_0(T_n^{(0)} > (1-\delta)a) + P_0(\hat{\gamma}_n^{-1} W_n > (1-\delta)a) + P_0([T_n^{(0)} > \delta a] \cap [\hat{\gamma}_n^{-1} W_n > \delta a]) \\ &= P_0(T_n^{(0)} > (1-\delta)a). \end{aligned}$$

Letting $\delta \downarrow 0$, we obtain $P_0(T_n^{(1)} > a) \leq P_0(T_n^{(0)} > a)$ for sufficiently large a. On the other hand,

$$P_0(T_n^{(1)} > a) \geq P_0(T_n^{(0)} > a + c, \ \hat{\gamma}_n^{-1} W_n > -c) = P_0(T_n^{(0)} > a + c).$$

Hence,

$$\frac{-\ln P_0(T_n^{(0)} > a)}{-\ln(1 - F(a))} \leq \frac{-\ln P_0(T_n^{(1)} > a)}{-\ln(1 - F(a))} \leq \frac{-\ln P_0(T_n^{(0)} > a + c)}{-\ln(1 - F(a))}$$

for sufficiently large a; it proves the theorem. □

Corollary 3.1. *(i) Let $T_n^{(0)} = \tilde{X}_n$ be the sample median, n odd. Let T_n be an equivariant estimator and $T_n^{(k)}$ its k-step version starting with $\tilde{X}_n$. Then, under the conditions of Theorem 3.2,*

$$\lim_{a \to \infty} B(T_n^{(k)}, a) = \frac{n+1}{2} \qquad \textit{for} \quad k = 1, 2, \dots.$$

(ii) Let $T_n^{(0)} = \bar{X}_n$ *be the sample mean. Let* T_n *be an equivariant estimator and* $T_n^{(k)}$ *its k-step version starting with* $\bar{X}_n$*. Then, under the conditions of Theorem 3.2, for* $k = 1, 2, \ldots,$

$$\lim_{a\to\infty} B(T_n^{(k)}, a) = \begin{cases} 1 & \text{if } F \text{ is exponentially tailed} \\ 0 & \text{if } F \text{ is heavy tailed.} \end{cases}$$

where the exponentially and heavy tailed F satisfy

$$\lim_{a\to\infty} \frac{-\ln(1-F(a))}{ba^r} = 1, \quad b > 0, \quad r \geq 1$$

$$\lim_{a\to\infty} \frac{-\ln(1-F(a))}{m\ln a} = 1, \quad m > 0,$$

respectively (see [2] for more details).

References

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

[2] J. Jurečková. Tail behavior of location estimators. *Ann. Statist.*, 9:578–585, (1981).

[3] J. Jurečková and M. Malý. The asymptotics for studentized k-step M-estimators of location. *Sequential Analysis*, 14:229–245, (1995).

[4] J. Jurečková and J. Picek. Minimum risk equivariant estimators in linear regression model. *Statistics & Decisions*, 27:37–59, (2009).

[5] J. Jurečková and J. Picek. Finite-sample behavior of robust estimators. In S. Chen et al., editor, *Recent Researches in Instrumentation, Measurement, Circuits and Systems*, pages 15–20. WSEAS, ISSN 1792-8575, (2011).

[6] J. Jurečková and S. Portnoy. Asymptotics for one-step M-estimators in regression with application to combining efficiency and high breakdown point. *Commun. Statist. A*, 16:2187–2199, (1987).

[7] J. Jurečková and Pranab K. Sen. Effect of the initial estimator on the asymptotic behavior of one-step M-estimator. *Ann. Inst. Statist. Math.*, 42:345–357, (1990).

[8] I. Mizera and C. H. Mueller. Breakdown points and variation exponents of robust M-estimators in linear models. *Ann. Statist*, 27:1164–1177, (1999).