NANZAN-TR-2001-02

The maximum asymptotic bias of S-estimates for regression over the neighborhoods defined by certain special capacities

Masakazu Ando and Miyoshi Kimura

January 2002

Technical Report of the Nanzan Academic Society
Mathematical Sciences and Information Engineering

THE MAXIMUM ASYMPTOTIC BIAS OF S-ESTIMATES FOR REGRESSION OVER THE NEIGHBORHOODS DEFINED BY CERTAIN SPECIAL CAPACITIES

By
Masakazu Ando and Miyoshi Kimura
Nanzan University

Abstract

The maximum asymptotic bias of an S-estimate for regression in the linear model is evaluated over the neighborhoods defined by certain special capacities, and its lower and upper bounds are derived. As special cases, the neighborhoods include those in terms of ε-contamination, total variation distance and Rieder's (ε, δ)-contamination. It is shown that when the model distribution is normal and the (ε, δ)-contamination neighborhood is adopted, the lower and upper bounds of an S-estimate (including the LMS-estimate) based on a jump function coincide with the maximum asymptotic bias. The tables of the maximum asymptotic bias of the LMS-estimate are given. These results are an extension of the corresponding ones due to Martin, Yohai and Zamar (1989), who used ε-contamination neighborhoods.

1. Introduction.

In the linear regression model the least squares (LS-) estimate of regression has been commonly used and in fact when the errors are distributed with a normal distribution, it minimizes the mean squared error in the class of all unbiased estimates. However, it is also well known that the LS-estimate is very sensitive to slight departures from normality or to the presence of a small proportion of outliers in the sample. Therefore, in the situations that the model is only approximately satisfied or some outliers may occur, it is desirable to use so-called robust estimates, which are not so sensitive to such departures or outliers and do not lose good properties so much. Various robust regression estimates have been proposed to date. As typical robust regression estimates, there are M-estimates (Huber,1973), generalized M- (GM-) estimates (Hampel et al., 1986), the least median of squares (LMS-) estimate (Rousseeuw, 1984), the least trimmed squares (LTS-) estimate (Rousseeuw, 1985), S-estimates (Rousseeuw and Yohai, 1984), MM-estimates (Yohai,1987), τ-estimates (Yohai and Zamar,1988) generalized S-(GS-) estimates (Hössjer, Croux and Rousseeuw, 1994), the least α-quantile (LαQ-) estimates, the least trimmed median (LTM-) estimate (Croux, Rousseeuw and Bael,1996) and so on.

Among them, S-estimates were introduced as high breakdown estimates which share the flexibility and nice asymptotic properties M-estimates have. S-estimates belong to the class of M-estimates with general scale and they include the LMS-estimate as an important special

AMS 1991 Subject classifications: Primary 62F35, Secondary 62G35, 62J05
Key words: S-estimate, LMS-estimate, robust regression, maximum asymptotic bias, ε-contamination, total variation, special capacity.

case. Various properties of S-estimates for regression were studied by Martin, Yohai and Zamar (1989), Davies (1990,1993), Hössjer (1992), He and Simpson (1993), Henning (1995), Berrendero and Zamar (2001) and others.

The most informative global quantitative measure to assese robustness of an estimate is the maximum asymptotic bias of the estimate caused by deviation from the assumed model distribution. It shows the whole preformance of the estimate from the model (i.e. no departure from the model) to the breakdown point, and it gives the gross error sensitivity as a local robustness measure, the breakdown point as a global robustness measure and a lot more. Martin, Yohai and Zamar (1989) derived the maximum asymptotic bias of an S-estimate over the ε-contamination neighborhood and showed that the minimax-bias M-estimate, which minimizes the maximum asymptotic bias in the class of all M-estimates with general scale, is given by an S-estimate based on a jump function. Although they adopted ε-contamination as deviation from the model, it is also valuable to evaluate the maximum asymptotic bias of an S-estimate over neighborhoods other than those in terms of ε-contamination.

The purpose of this paper is to derive lower and upper bounds on the maximum asymptotic bias of an S-estimate over the neighborhoods defined by certain special capacities. The neighborhoods were proposed and characterized by Ando and Kimura (2001). As special cases they include the neighborhoods in terms of ε-contamination, total variation and Rieder's (1977) (ε, δ)-contamination. It is shown that when the model distribution is normal and the (ε, δ)-contamination neighborhood is adopted, the derived lower and upper bounds for an S-estimate (including the LMS-estimate) based on a jump function coincide with the maximum asymptotic bias. In the case of ε- contamination neighborhoods, the lower and upper bounds reduce to the maximum asymptotic bias due to Martin, Yohai and Zamar (1989). Therefore our results give an extension of theirs.

The paper is organized as follows. Section 2 gives the definitions of the linear regression model, S-estimates and the neighborhoods generated by special capacities. It also presents a characterization theorem of the neighborhoods due to Ando and Kimura (2001), which is used throughout this paper. Section 3 derives the lower and upper bounds for an S-estimate, following two lemmas. It also proves that the lower and upper bounds for the maximum asymptotic bias of an S-estimate (including the LMS-estimate) based on a jump function are obtained as the limits of those for S-estimates based on strictly monotone functions. Section 4 considers the case that the model distribution is normal with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{I}$, the identity matrix, and shows that when the (ε, δ)-contamination neighborhood is adopted, the lower and upper bounds for an S-estimate based on a jump function are identical to the maximum asymptotic bias. It also gives some tables of the maximum asymptotic bias. All the proofs of the lemmas and theorems in this paper are collected in the final Section 5.

2. Preliminaries.

We consider the linear model

$$y = \boldsymbol{\theta}_0'\boldsymbol{x} + u, \tag{2.1}$$

where $\boldsymbol{x} = (x_1, \ldots, x_p)'$ is a random vector in R^p, $\boldsymbol{\theta}_0 = (\theta_{10}, \ldots, \theta_{p0})'$ is the vector in R^p of the true regression parameters and the error u is a random variable independent of $\boldsymbol{x}$. Let F_0 be the nominal distribution functions of u. Then the nominal distribution function H_0 of $(y, \boldsymbol{x})$

is

$$H_0(y, \boldsymbol{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} F_0(y - \boldsymbol{\theta}_0' \boldsymbol{s}) dG_0(\boldsymbol{s}). \tag{2.2}$$

We assume that G_0 is elliptical about the origin with a scatter matrix $\boldsymbol{A}$. Let $\mathcal{M}$ be the set of all probability distributions H on $(R^{p+1}, \mathcal{B}^{p+1})$, where $\mathcal{B}^{p+1}$ is the Borel σ-field on R^{p+1}. Let $\boldsymbol{T}$ be an R^p-valued functional defined on $\mathcal{M}$. Given a sample $(y_1, \boldsymbol{x}_1), \ldots, (y_n, \boldsymbol{x}_n)$ of size n from H, we define the corresponding estimate of $\boldsymbol{\theta}_0$ as $\boldsymbol{T}(H_n)$, where H_n is the empirical distribution of the sample. We assume that $\boldsymbol{T}$ is regression equivariant, i.e. if $\tilde{y} = y + \boldsymbol{x}'\boldsymbol{b}$ and $\tilde{\boldsymbol{x}} = \boldsymbol{C}'\boldsymbol{x}$ for some full rank $p \times p$ matrix $\boldsymbol{C}$, then $\boldsymbol{T}(\tilde{H}) = \boldsymbol{C}^{-1}[\boldsymbol{T}(H) + \boldsymbol{b}]$, where $\tilde{H}$ is the distribution of $(\tilde{y}, \tilde{\boldsymbol{x}})$. In this case, the model parameter is transformed to $\tilde{\boldsymbol{\theta}}_0 = \boldsymbol{C}^{-1}[\boldsymbol{\theta}_0 + \boldsymbol{b}]$.

The asymptotic bias of $\boldsymbol{T}$ at H is defined by

$$b_{\boldsymbol{A}}(\boldsymbol{T}, H) = [(\boldsymbol{T}(H) - \boldsymbol{\theta}_0)'\boldsymbol{A}(\boldsymbol{T}(H) - \boldsymbol{\theta}_0)]^{\frac{1}{2}}, \tag{2.3}$$

where $\boldsymbol{A} = \boldsymbol{A}(G_0)$ is an affine equivariant covariance functional of $\boldsymbol{x}$, i.e., if $\boldsymbol{x} \sim G_0$ and $\tilde{\boldsymbol{x}} = \boldsymbol{B}\boldsymbol{x}$ for some nonsingular $p \times p$ matrix $\boldsymbol{B}$, then $\boldsymbol{A}(\tilde{G}_0) = \boldsymbol{B}\boldsymbol{A}(G_0)\boldsymbol{B}'$. Since we only work with regression and affine equivariant estimates and $b_{\boldsymbol{A}}(\boldsymbol{T}, H)$ is invariant under regression and affine equivariant transformations, we can assume without loss of generality that $\boldsymbol{A} = \boldsymbol{I}$(the identity matrix),i.e., G_0 is spherical, and $\boldsymbol{\theta}_0 = \boldsymbol{0}$. Therefore the nominal model (2.2) becomes

$$H_0(y, \boldsymbol{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} F_0(y) dG_0(\|\boldsymbol{s}\|) \tag{2.4}$$

and the asymptotic bias (2.3) reduces to

$$b(\boldsymbol{T}, H) = \|\boldsymbol{T}(H)\|, \tag{2.5}$$

where $\|\cdot\|$ denotes the Euclidean norm. We assume that $\boldsymbol{T}$ is asymptotically unbiased at H_0 i.e., $\boldsymbol{T}(H_0) = \boldsymbol{0}$.

Let ρ be a real valued function defined on the real line R. We assume the following conditions of ρ, F_0 and G_0 (see Martin, Yohai and Zamar,1989).

A1. (i) ρ is symmetric and strictly increasing on $(0, \infty)$ with $\rho(0) = 0$.

(ii) ρ is bounded with $\lim_{u \to \infty} \rho(u) = 1$.

(iii) ρ has only a finite number of discontinuities.

A2. F_0 is absolutely continuous with density f_0 which is symmetric, continuous and strictly decreasing for $u > 0$.

A3. G_0 is spherical and $P_{G_0}(\boldsymbol{\theta}'\boldsymbol{x} = 0) = 0$ $\forall \boldsymbol{\theta} \in R^p$ with $\boldsymbol{\theta} \neq \boldsymbol{0}$.

Let $0 < b < 1$. Then, given a distribution F the M-scale functional is defined by

$$S(F) = \inf \left\{ s > 0 : E_F \left[\rho\left(\frac{u}{s}\right) \right] \leq b \right\}. \tag{2.6}$$

Given a sample $u_1, \ldots, u_n$ from F, the corresponding M-estimate of scale is $S(F_n)$ where F_n is the empirical distribution of $u_1, \ldots, u_n$. For any $\boldsymbol{\theta} \in R^p$ let $F^{\boldsymbol{\theta}}$ be the distribution of the residual

$$r(\boldsymbol{\theta}) = y - \boldsymbol{\theta}'\boldsymbol{x}.$$

To emphasize the independent roles of $\boldsymbol{\theta}$ and H, we let $S(\boldsymbol{\theta}, H) = S(F^{\theta})$.

A functional $\boldsymbol{T}(H)$ is said to be an S-estimate functional of regression if there exists a sequence $\{\boldsymbol{\theta}_n\} \subset R^p$ such that

$$\lim_{n\to\infty} \boldsymbol{\theta}_n = \boldsymbol{T}(H) \tag{2.7}$$

and

$$\lim_{n\to\infty} S(\boldsymbol{\theta}_n, H) = \inf_{\theta \in R^p} S(\boldsymbol{\theta}, H). \tag{2.8}$$

We note that S-estimates of regression are regression equivariant, and under very mild conditions they are consistent and asymptotically normal.

In order to express deviation from the nominal distribution H_0, we use the following neighborhood introduced by Ando and Kimura (2001).

$$\mathcal{P}_{H_0}(c,\ \gamma) = \{H \in \mathcal{M}\ :\ H(B) \le c\, H_0(B) + \gamma,\ {}^{\forall}B \in \mathcal{B}^{p+1}\}, \tag{2.9}$$

where $0 \le \gamma < 1$ and $1 - \gamma \le c < \infty$. This neighborhood (2.9) is regarded as a generalization of the (ε, δ)- contamination neighborhood $\mathcal{P}_{H_0}(1-\varepsilon,\ \varepsilon+\delta)$ introduced by Rieder (1977), and as special cases it includes the ε-contamination neighborhood $\mathcal{P}_{H_0}(1-\varepsilon,\ \varepsilon)$ and the total variation neighborhood $\mathcal{P}_{H_0}(1,\ \delta)$. The neighborhood $\mathcal{P}_{H_0}(c, \gamma)$ is characterized as follows:

Proposition 2.1 (Ando and Kimura, 2001) *For $0 \le \gamma < 1$ and $1 - \gamma \le c < \infty$ it holds that*

$$\mathcal{P}_{H_0}(c,\ \gamma) = \{H = c\,(H_0 - W) + \gamma\, K\ :\ W \in \mathcal{W}_{H_0,\lambda}, K \in \mathcal{M}\},$$

where $\mathcal{W}_{H_0,\lambda}$ is the set of all measures W such that $W(B) \le H_0(B)$ holds for ${}^{\forall}B \in \mathcal{B}^{p+1}$ and $W(R^{p+1}) = \lambda = (c + \gamma - 1)/c$.

The maximum asymptotic bias of an estimate $\boldsymbol{T}$ over $\mathcal{P}_{H_0}(c,\ \gamma)$ is defined as

$$B_{\boldsymbol{T}}(c,\ \gamma) = \sup\{\|\boldsymbol{T}(H)\|\ :\ H \in \mathcal{P}_{H_0}(c,\ \gamma)\}.$$

Martin, Yohai and Zamar (1989, Theorems 3.1 and 4.2) derived $B_{\boldsymbol{T}}(1-\varepsilon,\ \varepsilon)$ of an S-estimate $\boldsymbol{T}$ over $\mathcal{P}_{H_0}(1-\varepsilon,\ \varepsilon)$, and showed that the minimax-bias estimate, which minimizes $B_{\boldsymbol{T}}(1-\varepsilon,\ \varepsilon)$ in the class of all M-estimates with general scale, is given by an S-estimate based on a jump function ρ and some b.

3. The lower and upper bounds on the maximum asymptotic bias.

Let $\xi = \{W_{s,\theta}\ :\ s > 0, \boldsymbol{\theta} \in R^p\}$ be the set of $W_{s,\theta} \in \mathcal{W}_{H_0,\lambda}$ such that the (substochastic) distribution of $\frac{y - \boldsymbol{\theta}'\boldsymbol{x}}{s}$ under $H_0 - W_{s,\theta}$ depends on $\boldsymbol{\theta}$ only through $\|\boldsymbol{\theta}\|$, and let $\mathcal{F}_\lambda$ be the set of all such $\xi = \{W_{s,\theta}\}$. For any $\xi = \{W_{s,\theta}\} \in \mathcal{F}_\lambda$ we define a function g_ξ of s and $\|\boldsymbol{\theta}\|$ by

$$g_\xi(s, \|\boldsymbol{\theta}\|) = E_{H_0 - W_{s,\theta}}\left[\rho\left(\frac{y - \boldsymbol{\theta}'\boldsymbol{x}}{s}\right)\right]. \tag{3.1}$$

We consider the following condition of g_ξ:

A4. $g_\xi(s, \|\boldsymbol{\theta}\|)$ is continuous, strictly decreasing in s and strictly increasing in $\|\boldsymbol{\theta}\|$.

In order to establish an upper bound for $B_{\boldsymbol{T}}(c,\gamma)$ we need $\hat{\xi} = \{\hat{W}_{s,\theta}\}$ and $\xi^* = \{W^*_{s,\theta}\}$ defined as follows:

$$
(3.2) \quad
\begin{aligned}
\hat{W}_{s,\theta}(B) &= H_0\left(B \cap \left\{\left|\frac{y-\boldsymbol{\theta}'\boldsymbol{x}}{s}\right| \geq a_{s,\|\theta\|}\left(\frac{c+\gamma-1}{c}\right)\right\}\right), \quad {}^\forall B \in \mathcal{B}^{p+1},\\
W^*_{s,\theta}(B) &= H_0\left(B \cap \left\{\left|\frac{y-\boldsymbol{\theta}'\boldsymbol{x}}{s}\right| \leq a_{s,\|\theta\|}\left(\frac{1-\gamma}{c}\right)\right\}\right), \quad {}^\forall B \in \mathcal{B}^{p+1},
\end{aligned}
$$

where $a_{s,\|\theta\|}(\eta)$ $(0 < \eta < 1)$ denotes the upper $100\eta\%$ point of the distribution of $|y - \boldsymbol{\theta}'\boldsymbol{x}|/s$ under H_0 such that

$$
H_0\left(\left|\frac{y-\boldsymbol{\theta}'\boldsymbol{x}}{s}\right| \geq a_{s,\|\theta\|}(\eta)\right) \; = \; \eta.
$$

It is clear that $\hat{\xi} \in \mathcal{F}_\lambda$ and $\xi^* \in \mathcal{F}_\lambda$ follow from the definition (3.2) and A3. The following lemma shows that $g_{\hat{\xi}}$ and g_{ξ^*} satisfy A4.

Lemma 3.1 *Assume that* A1, A2 *and* A3 *are satisfied. Then* $g_{\hat{\xi}}$ *and* g_{ξ^*} *are continuous, strictly decreasing in* s *and strictly increasing in* $\|\boldsymbol{\theta}\|$.

We also obtain the following lemma.

Lemma 3.2 *Assume that* A1, A2 *and* A3 *are satisfied. Then for any* $s > 0$ *and any* $\boldsymbol{\theta} \in R^p$ *the following results hold.*

(i) $$\inf_{H \in \mathcal{P}_{H_0}(c,\gamma)} E_H\left[\rho\left(\frac{y-\boldsymbol{\theta}'\boldsymbol{x}}{s}\right)\right] = c\; g_{\hat{\xi}}(s, \|\boldsymbol{\theta}\|),$$

(ii) $$\sup_{H \in \mathcal{P}_{H_0}(c,\gamma)} E_H\left[\rho\left(\frac{y-\boldsymbol{\theta}'\boldsymbol{x}}{s}\right)\right] = c\; g_{\xi^*}(s, \|\boldsymbol{\theta}\|) + \gamma.$$

Let $\mathcal{F}_{0\lambda}$ be the set of all $\xi \in \mathcal{F}_\lambda$ satisfying A4. We note that Lemma 3.1 shows $\hat{\xi} \in \mathcal{F}_{0\lambda}$ and $\xi^* \in \mathcal{F}_{0\lambda}$. For any $\xi \in \mathcal{F}_{0\lambda}$ let $g^{-1}_{\xi,1}(\,\cdot\,, \|\boldsymbol{\theta}\|)$ be the inverse of g_ξ with respect to s and $g^{-1}_{\xi,2}(s,\,\cdot\,)$ the inverse of g_ξ with respect to $\|\boldsymbol{\theta}\|$. The following theorem corresponding to Theorem 3.1 of Martin, Yohai and Zamar (1989) gives the lower and upper bounds on $B_{\boldsymbol{T}}(c,\gamma)$ of an S-estimate $\boldsymbol{T}$ over $\mathcal{P}_{H_0}(c,\gamma)$.

Theorem 3.1 *Assume that* A1,A2 *and* A3 *are satisfied. Then*

$$
\begin{aligned}
&\underline{B}_{\boldsymbol{T}}(c,\gamma) \leq B_{\boldsymbol{T}}(c,\gamma) \leq \overline{B}_{\boldsymbol{T}}(c,\gamma), && \textit{if} \quad \gamma < \min(b, 1-b),\\
&B_{\boldsymbol{T}}(c,\gamma) = \infty, && \textit{if} \quad \gamma \geq \min(b, 1-b),
\end{aligned}
$$

where

$$\overline{B}_{\boldsymbol{T}}(c,\,\gamma) \;=\; g_{\hat{\xi},2}^{-1}\left(g_{\xi^*,1}^{-1}\left(\frac{b-\gamma}{c},0\right),\frac{b}{c}\right),$$
$$\underline{B}_{\boldsymbol{T}}(c,\,\gamma) \;=\; \sup_{\xi\in\mathcal{F}_{0\lambda}} g_{\xi,2}^{-1}\left(g_{\xi,1}^{-1}\left(\frac{b-\gamma}{c},0\right),\frac{b}{c}\right).$$

Remark 3.1 When $c = 1-\varepsilon$ and $\gamma = \varepsilon$ (i.e., the ε-contamination case), Theorem 3.1 reduces to Theorem 3.1 of Martin, Yohai and Zamar (1989). We note that ρ is assumed to be strictly increasing on $(0,\infty)$ instead of being nondecreasing.

We wish to show that Theorem 3.1 also holds for S-estimates (including the LMS-estimate) based on some type of step functions. To this end we need the following condition of ρ.

A5. ρ is symmetric and right continuous on $[0,\infty)$.

The following lemma enables us to achieve the purpose.

Lemma 3.3 *Let $\{\rho_n\}$ be a sequence of functions satisfying* A1 *and* A5*, and suppose that $\{\rho_n\}$ uniformly converges to a function $\tilde{\rho}$ on $(-\infty,\infty)$. Let $\boldsymbol{T}_n$ and $\tilde{\boldsymbol{T}}$ be the S-estimates based on ρ_n and $\tilde{\rho}$, respectively. Then for any $H\in\mathcal{P}_{H_0}(c,\gamma)$ there exists a subsequence $\{\boldsymbol{T}_{n_k}(H)\}$ of $\{\boldsymbol{T}_n(H)\}$ such that $\lim_{k\to\infty}\boldsymbol{T}_{n_k}(H) = \tilde{\boldsymbol{T}}(H)$.*

For any $\xi = \{W_{s,\theta}\}\in\mathcal{F}_{0\lambda}$ we let $g_{n,\xi}(s,\|\boldsymbol{\theta}\|)$, $g_{n,\xi,1}^{-1}(\,\cdot\,,\|\boldsymbol{\theta}\|)$ and $g_{n,\xi,2}^{-1}(s,\,\cdot\,)$ be $g_\xi(s,\|\boldsymbol{\theta}\|)$, $g_{\xi,1}^{-1}(\,\cdot\,,\|\boldsymbol{\theta}\|)$ and $g_{\xi,2}^{-1}(s,\,\cdot\,)$ based on ρ_n, respectively. Lemma 3.3 states that the following corollary holds.

Corollary 3.1 *Assume that all the conditions of Lemma 3.3 are satisfied. Then*

$$\begin{array}{ll} \underline{B}_{\tilde{\boldsymbol{T}}}(c,\gamma)\le B_{\tilde{\boldsymbol{T}}}(c,\gamma)\le \overline{B}_{\tilde{\boldsymbol{T}}}(c,\gamma), & \text{if } \gamma<\min(b,1-b),\\ B_{\tilde{\boldsymbol{T}}}(c,\gamma)=\infty, & \text{if } \gamma\ge\min(b,1-b), \end{array}$$

where

$$\overline{B}_{\tilde{\boldsymbol{T}}}(c,\gamma) \;=\; \lim_{k\to\infty} g_{n_k,\hat{\xi},2}^{-1}\left(g_{n_k,\xi^*,1}^{-1}\left(\frac{b-\gamma}{c},0\right),\frac{b}{c}\right),$$
$$\underline{B}_{\tilde{\boldsymbol{T}}}(c,\gamma) \;=\; \sup_{\xi\in\mathcal{F}_{0\lambda}}\left[\lim_{k\to\infty} g_{n_k,\xi,2}^{-1}\left(g_{n_k,\xi,1}^{-1}\left(\frac{b-\gamma}{c},0\right),\frac{b}{c}\right)\right].$$

4. The normal distribution model.

We consider the case that the nominal distribution H_0 of $(y, \boldsymbol{x})$ is the multivariate normal distribution $N(\boldsymbol{0}, \boldsymbol{I}_{p+1})$, where $\boldsymbol{I}_{p+1}$ denotes the $(p+1) \times (p+1)$ identity matrix. In this case, under H_0 $\frac{y-\boldsymbol{\theta}'\boldsymbol{x}}{s}$ is normally distributed with $N\left(0, \frac{1+\|\boldsymbol{\theta}\|^2}{s^2}\right)$. We denote by ϕ the density function of the standard normal distribution $N(0,1)$.

Let $\mathcal{F}_{1\lambda}$ be the set of $\xi = \{W_{s,\theta}\} \in \mathcal{F}_{0\lambda}$ such that the density function $f^{\xi}_{s,\|\theta\|}$ of $\frac{y-\boldsymbol{\theta}'\boldsymbol{x}}{s}$ under $H_0 - W_{s,\theta}$ is written in the form of

$$f^{\xi}_{s,\|\theta\|}(u) = \frac{s}{\sqrt{1+\|\boldsymbol{\theta}\|^2}}\phi_{\xi}\left(\frac{s}{\sqrt{1+\|\boldsymbol{\theta}\|^2}}u\right), \quad -\infty < u < \infty, \tag{4.1}$$

where ϕ_ξ is a function which does not depend on s and $\|\boldsymbol{\theta}\|$, and it satisfies $0 \le \phi_\xi \le \phi$ and

$$\int_{-\infty}^{\infty} \phi_\xi(u)du = \frac{1-\gamma}{c}. \tag{4.2}$$

It is easy to see that $\hat{\xi} = \{\hat{W}_{s,\theta}\}$ and $\xi^* = \{W^*_{s,\theta}\}$ are elements of $\mathcal{F}_{1\lambda}$ with

$$\phi_{\hat{\xi}}(u) \;=\; \begin{cases} \phi(u) & \text{if } \; 0 \le |u| \le z_{\left(\frac{c+\gamma-1}{2c}\right)}, \\ 0 & \text{if } \; |u| > z_{\left(\frac{c+\gamma-1}{2c}\right)}, \end{cases} \tag{4.3}$$

and

$$\phi_{\xi^*}(u) \;=\; \begin{cases} 0 & \text{if } \; 0 \le |u| < z_{\left(\frac{1-\gamma}{2c}\right)}, \\ \phi(u) & \text{if } \; |u| \ge z_{\left(\frac{1-\gamma}{2c}\right)}, \end{cases} \tag{4.4}$$

respectively, where $z_{(\eta)}$ denotes the upper $100\eta\%$ point of $N(0,1)$.

For any $\xi = \{W_{s,\theta}\} \in \mathcal{F}_{1\lambda}$ and any $\tau > 0$ we define

$$h_\xi(\tau) = \int_{-\infty}^{\infty} \rho(u)\frac{1}{\tau}\phi_\xi\left(\frac{1}{\tau}u\right)du. \tag{4.5}$$

Then we can see

$$g_\xi(s, \|\boldsymbol{\theta}\|) = h_\xi(\sqrt{1+\|\boldsymbol{\theta}\|^2}/s), \tag{4.6}$$

and hence

$$\begin{aligned} g^{-1}_{\xi,1}(t, \|\boldsymbol{\theta}\|) &= \sqrt{1+\|\boldsymbol{\theta}\|^2}/h^{-1}_\xi(t), \\ g^{-1}_{\xi,2}(s,t) &= \sqrt{(sh^{-1}_\xi(t))^2 - 1}. \end{aligned} \tag{4.7}$$

Thus we obtain the following corollary from Theorem 3.1.

Corollary 4.1 *Assume that the nominal distribution H_0 of $(y, \boldsymbol{x})$ is $N(\mathbf{0}, \boldsymbol{I}_{p+1})$ and that* A1 *is satisfied. Then*

$$\sup_{\xi\in\mathcal{F}_{1\lambda}} \left[\frac{h_\xi^{-1}\left(\frac{b}{c}\right)}{h_\xi^{-1}\left(\frac{b-\gamma}{c}\right)}\right]^2 - 1 \le B^2_{\boldsymbol{T}}(c,\gamma) \le \left[\frac{h_{\hat\xi}^{-1}\left(\frac{b}{c}\right)}{h_{\xi^*}^{-1}\left(\frac{b-\gamma}{c}\right)}\right]^2 - 1. \tag{4.8}$$

Next we consider S-estimates based on the following jump function:

$$\tilde\rho_\alpha(u) = \begin{cases} 0 & \text{if } |u| < \alpha, \\ 1 & \text{if } |u| \ge \alpha. \end{cases} \tag{4.9}$$

We note that $\tilde\rho_\alpha$ does not satisfy the condition (i) of A1. Since the S-estimate based on $\tilde\rho_\alpha$ does not depend on the choice of α, we hereafter take $\alpha = \alpha_0 = \Phi^{-1}(\frac{3}{4})$ and denote $\tilde\rho = \tilde\rho_{\alpha_0}$ for simplicity. For a given sample $\boldsymbol{u} = (u_1, \cdots, u_n)$, the M-estimate of scale based on $\tilde\rho$ is given by

$$s_n(\boldsymbol{u}) = \frac{1}{\alpha_0}|u|_{(n-[nb])},$$

where $|u|_{(1)}, \cdots, |u|_{(n)}$ are the order statistics for $|u_1|, \cdots, |u_n|$, and $[a]$ denotes the largest integer smaller than or equal to a. The S-estimate based on $\tilde\rho$ and b is denoted by $\boldsymbol{T}_b$.

Note that $\boldsymbol{T}_b (= \boldsymbol{T}_{0.5})$ corresponding to $b = 0.5$ is Rousseeuw's (1984) least median of squares (LMS-) estimate. Martin, Yohai and Zamar (1989) shows that in the case of ε-contamination neighborhoods the estimate $\boldsymbol{T}_b$ with a properly chosen b minimizes $B_{\boldsymbol{T}}(1-\varepsilon, \varepsilon)$ in the class of M-estimates with general scale.

We are now interested in evaluating $B_{\boldsymbol{T}_b}(c,\gamma)$. Although $\tilde\rho$ does not satisfy the condition (i) of A1, there exists a sequence $\{\rho_n\}$ of functions satisfying A1 and A5 such that $\{\rho_n\}$ converges to $\tilde\rho$ uniformly on $(-\infty, \infty)$. For example we can take ρ_n as follows:

$$\rho_n(u) = \begin{cases} \left(\dfrac{|u|}{(n+1)\alpha_0}\right), & \text{if } |u| < \alpha_0, \\ 1 - \dfrac{2\alpha_0^2 e^{\frac{1}{2}\alpha_0^2}}{(n+1)(1+2\alpha_0^2)}\left(1 + \dfrac{1}{2u^2}\right)e^{-\frac{1}{2}u^2}, & \text{if } |u| \ge \alpha_0. \end{cases} \tag{4.10}$$

For any $\xi = \{W_{s,\theta}\} \in \mathcal{F}_{1\lambda}$ let $h_{\rho_n,\xi}$ and $h_{\tilde\rho,\xi}$ denote h_ξ with ρ replaced by ρ_n and $\tilde\rho$ in (4.5), respectively. Then for any ξ we have

$$\begin{aligned} &\lim_{n\to\infty} h_{\rho_n,\xi}(\tau) = h_{\tilde\rho,\xi}(\tau), \quad {}^\forall\tau > 0, \\ &\lim_{n\to\infty} h^{-1}_{\rho_n,\xi}(t) = h^{-1}_{\tilde\rho,\xi}(t), \quad 0 < {}^\forall t < \tfrac{1-\gamma}{c}. \end{aligned} \tag{4.11}$$

As easily seen, we also have

$$h_{\tilde\rho,\hat\xi}(\tau) = \begin{cases} 0, & \text{if } 0 \le \tau \le \frac{1}{z_{(\frac{c+\gamma-1}{2c})}}, \\ 2\left[1 - \frac{c+\gamma-1}{2c} - \Phi\left(\frac{1}{\tau}\right)\right], & \text{if } \tau > \frac{1}{z_{(\frac{c+\gamma-1}{2c})}}, \end{cases} \tag{4.12}$$

and

$$(4.13) \qquad h_{\tilde{\rho},\xi^*}(\lambda) = \begin{cases} 2\left[1-\Phi\left(\frac{1}{\tau}\right)\right], & \text{if } \ 0 \le \tau \le \frac{1}{z_{\left(\frac{1-\gamma}{2c}\right)}}, \\ \frac{1-\gamma}{c}, & \text{if } \ \tau \ge \frac{1}{z_{\left(\frac{1-\gamma}{2c}\right)}}, \end{cases}.$$

Therefore we can obtain the following corollary from Corollaries 3.1 and 4.1.

Corollary 4.2 *Assume that the nominal distribution* H_0 *of* $(y, \boldsymbol{x})$ *is* $N(\mathbf{0}, \boldsymbol{I}_{p+1})$. *Then*

$$(4.14) \qquad \begin{aligned} &\underline{B}_{\boldsymbol{T}_b}(c,\gamma) \le B_{\boldsymbol{T}_b}(c,\gamma) \le \overline{B}_{\boldsymbol{T}_b}(c,\gamma), && \text{if } \ \gamma < \min(b, 1-b), \\ &B_{\boldsymbol{T}_b}(c,\gamma) = \infty, && \text{if } \ \gamma \ge \min(b, 1-b), \end{aligned}$$

where

$$\underline{B}^2_{\boldsymbol{T}_b}(c,\gamma) = \sup_{\xi\in\mathcal{F}_{1\lambda}} \left[\frac{h^{-1}_{\tilde{\rho},\xi}\left(\frac{b}{c}\right)}{h^{-1}_{\tilde{\rho},\xi}\left(\frac{b-\gamma}{c}\right)}\right]^2 - 1 \quad \text{and} \quad \overline{B}^2_{\boldsymbol{T}_b}(c,\gamma) = \left[\frac{h^{-1}_{\tilde{\rho},\hat{\xi}}\left(\frac{b}{c}\right)}{h^{-1}_{\tilde{\rho},\xi^*}\left(\frac{b-\gamma}{c}\right)}\right]^2 - 1.$$

In this Corollary 4.2 we see

$$(4.15) \qquad \frac{h^{-1}_{\tilde{\rho},\hat{\xi}}\left(\frac{b}{c}\right)}{h^{-1}_{\tilde{\rho},\xi^*}\left(\frac{b-\gamma}{c}\right)} = \frac{\Phi^{-1}\left(1-\frac{b-\gamma}{2c}\right)}{\Phi^{-1}\left(1-\frac{b+\gamma+c-1}{2c}\right)}.$$

On the other hand, as for the lower bound, let us consider $\xi = \{W_{s,\theta}\}$ such that $W_{s,\theta} = \lambda H_0$ for ${}^\forall s > 0$ and ${}^\forall \boldsymbol{\theta} \in R^p$. Then we have $H_0 - W_{s,\theta} = (1-\lambda)H_0$, and $\xi \in \mathcal{F}_{1\lambda}$ follows from Lemma 3.1 of Martin, Yohai and Zamar (1989).

Since

$$h_{\tilde{\rho},\xi}(\tau) = 2(1-\lambda)\left(1-\Phi\left(\frac{1}{\tau}\right)\right),$$

it follows that

$$(4.16) \qquad \frac{h^{-1}_{\tilde{\rho},\xi}\left(\frac{b}{c}\right)}{h^{-1}_{\tilde{\rho},\xi}\left(\frac{b-\gamma}{c}\right)} = \frac{\Phi^{-1}\left(1-\frac{b-\gamma}{2(1-\gamma)}\right)}{\Phi^{-1}\left(1-\frac{b}{2(1-\gamma)}\right)}.$$

Therefore

$$\underline{B}^2_{\boldsymbol{T}_b}(c,\gamma) \ge \left[\frac{\Phi^{-1}\left(1-\frac{b-\gamma}{2(1-\gamma)}\right)}{\Phi^{-1}\left(1-\frac{b}{2(1-\gamma)}\right)}\right]^2 - 1.$$

We note that the right hand side is equal to $B^2_{\boldsymbol{T}_b}(1-\gamma,\gamma)$. When $\lambda = 0$, we have $\underline{B}_{\boldsymbol{T}_b}(c,\gamma) = \overline{B}_{\boldsymbol{T}_b}(c,\gamma) = B_{\boldsymbol{T}_b}(c,\gamma) = B_{\boldsymbol{T}_b}(1-\gamma,\gamma)$.

The following theorem states that in Rieder's neighborhood case (i.e., $c < 1$), we can obtain $B_{\boldsymbol{T}_b}(c,\gamma)$ exactly.

Theorem 4.1 *Assume that the nominal distribution H_0 of $(y, \boldsymbol{x})$ is $N(\mathbf{0}, \boldsymbol{I}_{p+1})$ and let $c < 1$. Then*

$$(4.17) \qquad B^2_{\boldsymbol{T}_b}(c,\gamma) = \begin{cases} \left[\dfrac{\Phi^{-1}\left(1-\frac{b-\gamma}{2c}\right)}{\Phi^{-1}\left(1-\frac{b+\gamma+c-1}{2c}\right)}\right]^2 - 1, & \text{if } \gamma < \min(b, 1-b), \\ 0, & \text{if } \gamma \geq \min(b, 1-b). \end{cases}$$

Remark 4.1

1. When $1-\gamma \leq c < 1$, the neighborhood $\mathcal{P}_{H_0}(c,\gamma)$ is the one introduced by Rieder (1977) (let $c = 1-\varepsilon$ and $\gamma = \varepsilon + \delta$).

2. When $c = 1-\varepsilon$ and $\gamma = \varepsilon$, Theorem 4.1 becomes

$$B^2_{\boldsymbol{T}_b}(1-\varepsilon,\varepsilon) = \left[\frac{\Phi^{-1}\left(1-\frac{b-\varepsilon}{2(1-\varepsilon)}\right)}{\Phi^{-1}\left(1-\frac{b}{2(1-\varepsilon)}\right)}\right]^2 - 1,$$

which is the same as (3.24) in Martin, Yohai and Zamar (1989).

We are concerned with the LMS-estimate $\boldsymbol{T}_{0.5}$ with $b = 0.5$. Table 1 shows the values of $B_{\boldsymbol{T}_{0.5}}(c,\ \gamma)$ for selected $c < 1$ and γ. Note that the values on the diagonal line of the table correspond to the case of ε-contamination neighborhoods. Table 2 gives the upper bounds $\overline{B}_{\boldsymbol{T}_{0.5}}(c,\ \gamma)$ for selected $c \geq 1$ and γ. Table 3 presents the values of $B_{\boldsymbol{T}_{0.5}}(1-\varepsilon, \varepsilon+\delta)$ for selected ε and δ $(0 \leq \varepsilon < 0.5,\ \ 0 \leq \delta < 0.5)$, which shows the maximum asymptotic bias of $\boldsymbol{T}_{0.5}$ over Rieder's (ε,δ)-contamination neighborhoods.

Table 1: $B_{\boldsymbol{T}_b}(c,\ \gamma)$ for $b = 0.5$ $\ (c < 1,\ 0 \leq \gamma < 0.5)$

$c \setminus \gamma$	0.00	0.01	0.02	0.03	0.04	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
0.55	—	—	—	—	—	—	—	—	—	—	—	—	—	14.77
0.60	—	—	—	—	—	—	—	—	—	—	—	—	6.49	16.52
0.65	—	—	—	—	—	—	—	—	—	—	—	3.96	7.28	18.29
0.70	—	—	—	—	—	—	—	—	—	—	2.73	4.45	8.07	20.08
0.75	—	—	—	—	—	—	—	—	—	2.01	3.10	4.95	8.88	21.90
0.80	—	—	—	—	—	—	—	—	1.51	2.30	3.46	5.46	9.70	23.73
0.85	—	—	—	—	—	—	—	1.14	1.76	2.59	3.83	5.97	10.52	25.58
0.90	—	—	—	—	—	—	0.82	1.36	2.01	2.88	4.20	6.49	11.36	27.45
0.95	—	—	—	—	—	0.52	1.05	1.58	2.25	3.17	4.58	7.01	12.20	29.34
0.96	—	—	—	—	0.46	0.58	1.09	1.63	2.30	3.23	4.65	7.12	12.37	29.71
0.97	—	—	—	0.39	0.52	0.63	1.13	1.67	2.35	3.29	4.73	7.22	12.54	30.09
0.98	—	—	0.31	0.45	0.57	0.68	1.17	1.71	2.40	3.35	4.80	7.33	12.71	30.47
0.99	—	0.22	0.39	0.51	0.62	0.72	1.21	1.75	2.44	3.41	4.88	7.44	12.88	30.85
1.00	0.00	0.31	0.45	0.56	0.67	0.77	1.25	1.80	2.49	3.46	4.95	7.54	13.05	31.24

Table 2: $\overline{B}_{\boldsymbol{T}_b}(c,\ \gamma)$ for $b = 0.5$ $\ (c \geq 1,\ 0 \leq \gamma < 0.5)$

$c \setminus \gamma$	0.00	0.01	0.02	0.03	0.04	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
1.0	0.00	0.31	0.45	0.57	0.67	0.77	1.26	1.80	2.50	3.47	4.96	7.55	13.05	31.24
1.1	0.73	0.82	0.90	0.99	1.08	1.17	1.64	2.22	2.98	4.06	5.72	8.61	14.77	35.08
1.2	1.09	1.17	1.26	1.34	1.43	1.51	2.01	2.64	3.47	4.66	6.50	9.70	16.52	38.97
1.3	1.41	1.49	1.58	1.66	1.75	1.85	2.38	3.05	3.96	5.26	7.28	10.80	18.29	42.91
1.4	1.72	1.80	1.89	1.98	2.07	2.17	2.74	3.47	4.46	5.88	8.08	11.92	20.09	46.89
1.5	2.01	2.10	2.19	2.29	2.39	2.50	3.10	3.89	4.96	6.50	8.88	13.05	21.90	50.91
1.6	2.30	2.40	2.50	2.60	2.71	2.82	3.47	4.31	5.47	7.12	9.70	14.20	23.73	54.97
1.8	2.88	2.99	3.10	3.22	3.34	3.47	4.21	5.18	6.50	8.40	11.36	16.52	27.46	63.18
2.0	3.47	3.59	3.71	3.84	3.98	4.13	4.96	6.05	7.55	9.70	13.05	18.89	31.24	71.52
2.5	4.96	5.11	5.27	5.44	5.62	5.81	6.89	8.31	10.25	13.05	17.40	24.97	40.94	92.79
3.0	6.50	6.69	6.89	7.10	7.32	7.55	8.88	10.64	13.05	16.52	21.90	31.24	50.91	114.59
4.0	9.70	9.97	10.25	10.54	10.85	11.17	13.05	15.52	18.89	23.73	31.24	44.23	71.52	159.42
5.0	13.05	13.40	13.76	14.15	14.55	14.97	17.40	20.60	24.97	31.24	40.94	57.69	92.79	205.51
10.0	31.24	32.02	32.83	33.69	34.58	35.51	40.94	48.04	57.69	71.52	92.79	129.37	205.51	447.93

Table 3: $B_{\boldsymbol{T}_b}(1-\varepsilon,\ \varepsilon+\delta)$ for $b = 0.5$ $\ (0 \leq \varepsilon < 0.5,\ 0 \leq \gamma < 0.5)$

$\varepsilon \setminus \delta$	0.00	0.01	0.02	0.03	0.04	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
0.00	0.00	0.31	0.45	0.57	0.67	0.77	1.26	1.80	2.50	3.47	4.96	7.55	13.05	31.24
0.01	0.22	0.39	0.52	0.63	0.73	0.83	1.32	1.88	2.61	3.65	5.28	8.19	14.77	40.44
0.02	0.32	0.46	0.58	0.68	0.78	0.88	1.38	1.97	2.74	3.85	5.64	8.95	16.96	56.33
0.03	0.39	0.52	0.63	0.74	0.84	0.94	1.45	2.06	2.87	4.07	6.04	9.84	19.83	89.56
0.04	0.46	0.58	0.69	0.79	0.89	0.99	1.51	2.15	3.02	4.31	6.50	10.90	23.73	196.21
0.05	0.53	0.64	0.75	0.85	0.95	1.05	1.59	2.25	3.18	4.58	7.02	12.20	29.34	—
0.10	0.83	0.93	1.04	1.15	1.26	1.37	2.01	2.88	4.21	6.50	11.36	27.46	—	—
0.15	1.14	1.26	1.37	1.50	1.63	1.77	2.59	3.84	5.98	10.53	25.59	—	—	—
0.20	1.51	1.65	1.80	1.96	2.12	2.30	3.47	5.47	9.70	23.73	—	—	—	—
0.25	2.01	2.19	2.39	2.61	2.84	3.10	4.96	8.88	21.90	—	—	—	—	—
0.30	2.74	3.01	3.31	3.64	4.02	4.46	8.08	20.09	—	—	—	—	—	—
0.35	3.96	4.42	4.96	5.59	6.35	7.28	18.29	—	—	—	—	—	—	—
0.40	6.50	7.55	8.88	10.64	13.05	16.52	—	—	—	—	—	—	—	—
0.45	14.77	19.63	28.08	45.89	103.63	—	—	—	—	—	—	—	—	—

5. Proofs.

Proof of Lemma 3.1 Let $F_{s,\|\theta\|}$ and $F^{\xi}_{s,\|\theta\|}$ denote the distributions of $|\frac{y-\boldsymbol{\theta}'\boldsymbol{x}}{s}|$ under H_0 and $H_0 - W_{s,\theta}$, respectively, where $\xi = \{W_{s,\theta}\} \in \mathcal{F}_\lambda$. Then for $\hat{\xi} = \{\hat{W}_{s,\theta}\}$ and $\xi^* = \{W^*_{s,\theta}\}$ we can see that

$$F^{\hat{\xi}}_{s,\|\theta\|}(u) \;=\; \min\left(F_{s,\|\theta\|}(u), \frac{1-\gamma}{c}\right), \quad u \geq 0, \tag{5.1}$$

$$F^{\xi^*}_{s,\|\theta\|}(u) \;=\; \max\left(F_{s,\|\theta\|}(u)-\frac{c+\gamma-1}{c},0\right),\quad u\ge 0.$$

From Lemma 3.1 of Martin, Yohai and Zamar (1989) it follows that $F_{s,\|\theta\|}(u)$ is continuous in u, s and $\|\boldsymbol{\theta}\|$, strictly increasing in s and strictly decreasing in $\|\boldsymbol{\theta}\|$. Hence, (5.1) implies that $F^{\hat\xi}_{s,\|\theta\|}$ and $F^{\xi^*}_{s,\|\theta\|}$ also possess these properties of $F_{s,\|\theta\|}$. Therefore, noting that

$$\begin{aligned} g_{\hat\xi}(s,\|\boldsymbol{\theta}\|) &= E_{F^{\hat\xi}_{s,\|\theta\|}}[\rho(u)],\quad u\sim F^{\hat\xi}_{s,\|\theta\|},\\ g_{\xi^*}(s,\|\boldsymbol{\theta}\|) &= E_{F^{\xi^*}_{s,\|\theta\|}}[\rho(u)],\quad u\sim F^{\xi^*}_{s,\|\theta\|}, \end{aligned}\tag{5.2}$$

we obtain the lemma from A1. □

Proof of Lemma 3.2

(i) For $\hat\xi=\{\hat W_{s,\theta}\}$ let

$$H^{\hat\xi}_{s,\theta} \;=\; c\,(H_0-\hat W_{s,\theta})+\gamma\Delta_0,\tag{5.3}$$

where Δ_0 is the distribution with probability 1 at the origin. It is obvious that $H^{\hat\xi}_{s,\theta}\in\mathcal{P}_{H_0}(c,\gamma)$ holds. As easily seen, the distribution of $|\frac{y-\boldsymbol{\theta}'\boldsymbol{x}}{s}|$ under $H^{\hat\xi}_{s,\theta}$ is stochastically smaller than that under any $H\in\mathcal{P}_{H_0}(c,\gamma)$. Hence, the assertion (i) follows from A1 and the fact

$$E_{H^{\hat\xi}_{s,\theta}}\left[\rho\left(\frac{y-\boldsymbol{\theta}'\boldsymbol{x}}{s}\right)\right] \;=\; c\,E_{H_0-\hat W_{s,\theta}}\left[\rho\left(\frac{y-\boldsymbol{\theta}'\boldsymbol{x}}{s}\right)\right]+\gamma\,\rho(0)=c\,g_{\hat\xi}(s,\|\boldsymbol{\theta}\|).$$

(ii) For $\xi^*=\{W^*_{s,\theta}\}$ let

$$H^{\xi^*}_{n,s,\theta}=c\,(H_0-W^*_{s,\theta})+\gamma\Delta^*_n,\quad n=1,2,\dots,\tag{5.4}$$

where Δ^*_n is the distribution with probability 1 at $(y_n,\boldsymbol{x}_n)$ such that $\boldsymbol{x}_n=\lambda_n\boldsymbol{\theta}$, $y_n=2\lambda_n\|\boldsymbol{\theta}\|^2$ and $\lambda_n\to\infty$. Then it is clear that $H^{\xi^*}_{n,s,\theta}\in\mathcal{P}_{H_0}(c,\gamma)$. Since $|y_n-\boldsymbol{\theta}'\boldsymbol{x}_n|=\lambda_n\|\boldsymbol{\theta}\|\to\infty$, the distribution of $|\frac{y-\boldsymbol{\theta}'\boldsymbol{x}}{s}|$ under $H^{\xi^*}_{n,s,\theta}$ is stochastically larger than that under any $H\in\mathcal{P}_{H_0}(c,\gamma)$ for sufficiently large n. Hence the assertion (ii) follows from A1 and the fact

$$\begin{aligned}\lim_{n\to\infty}E_{H^{\xi^*}_{n,s,\theta}}\left[\rho\left(\frac{y-\boldsymbol{\theta}'\boldsymbol{x}}{s}\right)\right] &= c\,E_{H_0-W^*_{s,\theta}}\left[\rho\left(\frac{y-\boldsymbol{\theta}'\boldsymbol{x}}{s}\right)\right]+\gamma\lim_{n\to\infty}\rho\left(\frac{y_n-\boldsymbol{\theta}'\boldsymbol{x}_n}{s}\right)\\ &= c\,g_{\xi^*}(s,\|\boldsymbol{\theta}\|)+\gamma.\ \square\end{aligned}$$

Proof of Theorem 3.1 We first show $B_{\boldsymbol{T}}(c,\gamma)\le\overline{B}_{\boldsymbol{T}}(c,\gamma)$. Let $\bar d=\overline{B}_{\boldsymbol{T}}(c,\gamma)$. To complete the proof it is sufficient to show that for any $H=c(H_0-W)+\gamma K\in\mathcal{P}_{H_0}(c,\gamma)$ and any $\|\boldsymbol{\theta}\|>\bar d$ we have

$$s(\boldsymbol{\theta},H)\;>\;s(\mathbf{0},H),$$

because by the definition of $\boldsymbol{T}$ this implies $\boldsymbol{T}(H)\le\bar d$. Let

$$s^*\;=\;g^{-1}_{\xi^*,1}\left(\frac{b-\gamma}{c},0\right).$$

Since

$$c\, g_{\hat{\xi}}(s^*, \|\boldsymbol{\theta}\|) > c\, g_{\hat{\xi}}(s^*, \bar{d}) = b,$$

there exists $s_1 > s^*$ such that

$$c\, g_{\hat{\xi}}(s_1, \|\boldsymbol{\theta}\|) \quad > \quad b,$$

Hence, by (i) of Lemma 3.2

$$E_H\left[\rho\left(\frac{y - \boldsymbol{\theta}'\boldsymbol{x}}{s_1}\right)\right] \geq c\, g_{\hat{\xi}}(s_1, \|\boldsymbol{\theta}\|) > b.$$

This implies $s_1 \leq s(\boldsymbol{\theta}, H)$ and hence

$$s^* < s(\boldsymbol{\theta}, H). \tag{5.5}$$

On the other hand, it follows from (ii) of Lemma 3.2 that

$$E_H\left[\rho\left(\frac{y}{s^*}\right)\right] \leq c\, g_{\xi^*}(s^*, 0) + \gamma = b,$$

This implies

$$s^* \geq s(\mathbf{0}, H), \tag{5.6}$$

From (5.5) and (5.6) it follows that

$$s(\boldsymbol{\theta}, H) > s(\mathbf{0}, H).$$

Secondly, we show $B_{\boldsymbol{T}}(c,\, \gamma) \geq \underline{B}_{\boldsymbol{T}}(c,\, \gamma)$. To this end it is sufficient to show that $B^{\xi}_{\boldsymbol{T}}(c,\, \gamma) \leq B_{\boldsymbol{T}}(c,\, \gamma)$ holds for any $\xi \in \mathcal{F}_{0\lambda}$. Let $d_\xi = B^{\xi}_{\boldsymbol{T}}(c,\, \gamma)$, let d_1 be any positive real number smaller than d_ξ $(0 < d_1 < d_\xi)$ and let $\|\boldsymbol{\theta}^*\| = d_1$. Further, let

$$H_n \quad = \quad H^{\xi}_{n,s,\theta} = c\,(H_0 - W_{s,\theta}) + \gamma\Delta_n,$$

where Δ_n is the distribution with probability 1 at $(y_n, \boldsymbol{x}_n)$ with $\boldsymbol{x}_n = \lambda_n\boldsymbol{\theta}^*$, $\lambda_n \to \infty$ and $y_n = \boldsymbol{\theta}^{*\prime}\boldsymbol{x}_n$. To complete the proof it is sufficient to show

$$\sup_n \|\boldsymbol{T}(H_n)\| \quad \geq \quad d_1. \tag{5.7}$$

Assume that (5.7) is not true. Then there exists a subsequence $\{\boldsymbol{\theta}_{i_n}\}$ of $\{\boldsymbol{\theta}_n\}$ such that $\boldsymbol{T}(H_{i_n}) = \boldsymbol{\theta}_{i_n}$, $\lim_{n\to\infty} \boldsymbol{\theta}_{i_n} = \tilde{\boldsymbol{\theta}}$ and $\|\tilde{\boldsymbol{\theta}}\| < \|\boldsymbol{\theta}^*\| = d_1$. In what follows, we denote i_n by n for simplicity.

By Lemma 3.1, we have

$$E_{H_n}\left[\rho\left(\frac{y - \boldsymbol{\theta}_n'\boldsymbol{x}}{s}\right)\right] \quad > \quad c\, g_{\xi}(s, 0) + \gamma\rho\left(\frac{y_n - \boldsymbol{\theta}_n'\boldsymbol{x}_n}{s}\right).$$

Let $\tilde{s} = g^{-1}_{\xi,1}\left(\frac{b-\gamma}{c}, 0\right)$. Since

$$\lim_{n\to\infty} |y_n - \boldsymbol{\theta}_n'\boldsymbol{x}_n| = \lim_{n\to\infty} \lambda_n(\|\boldsymbol{\theta}^*\|^2 - \boldsymbol{\theta}^{*\prime}\boldsymbol{\theta}_n) = \infty,$$

we see that for any $s < \tilde{s}$

$$\lim_{n\to\infty} E_{H_n}\left[\rho\left(\frac{y-\boldsymbol{\theta}'_n\boldsymbol{x}}{s}\right)\right] \;\; > \;\; c\, g_\xi(\tilde{s}, 0) + \gamma = b.$$

This implies that

$$\lim_{n\to\infty} s(\boldsymbol{\theta}_n, H_n) \;\; \geq \;\; s, \quad {}^\forall s < \tilde{s}.$$

Thus

$$\lim_{n\to\infty} s(\boldsymbol{\theta}_n, H_n) \;\; \geq \;\; \tilde{s}. \tag{5.8}$$

On the other hand, we see

$$c\, g_\xi(\tilde{s}, d_1) < c\, g_\xi(\tilde{s}, d_\xi) = b,$$

and hence there exists $s_1 < \tilde{s}$ such that $c\, g_\xi(s_1, d_1) < b$. Hence

$$\begin{aligned} E_{H_n}\left[\rho\left(\frac{y-\boldsymbol{\theta}^{*\prime}\boldsymbol{x}}{s_1}\right)\right] \;\; &= \;\; c\, g_\xi(s_1, \|\boldsymbol{\theta}^*\|) + \gamma\rho\left(\frac{y_n-\boldsymbol{\theta}^{*\prime}\boldsymbol{x}_n}{s_1}\right) \\ &= \;\; c\, g_\xi(s_1, d_1) < b. \end{aligned}$$

This implies

$$s(\boldsymbol{\theta}^*, H_n) \;\; \leq \;\; s_1. \tag{5.9}$$

From (5.8) and (5.9) it follows that

$$s(\boldsymbol{\theta}^*, H_n) \leq s_1 < \tilde{s} \leq \lim_{n\to\infty} s(\boldsymbol{\theta}_n, H_n).$$

Since this contradicts the fact that $\boldsymbol{T}(H_n) = \theta_n$ minimizes $s(\;\cdot\;, H_n)$ for each n, we obtain (5.7) and hence

$$B^\xi_{\boldsymbol{T}}(c, \gamma) \;\; \leq \;\; B_{\boldsymbol{T}}(c, \gamma).$$

To show the last part of the theorem, we let $b \leq 0.5$. Then $\min(b, 1-b) = b$ and

$$\lim_{\gamma\uparrow b} g^{-1}_{\xi^*,1}\left(\frac{b-\gamma}{c}, 0\right) = \lim_{t\to 0} g^{-1}_{\xi^*,1}(t, 0) = \infty,$$

Hence

$$\lim_{\gamma\uparrow b} g^{-1}_{\hat{\xi},2}\left(g^{-1}_{\xi^*,1}\left(\frac{b-\gamma}{c}, 0\right), \frac{b}{c}\right) = \lim_{s\to\infty} g^{-1}_{\hat{\xi},2}\left(s, \frac{b}{c}\right) = \infty.$$

Similarly, we can show

$$\lim_{\gamma\uparrow b} g^{-1}_{\xi,2}\left(g^{-1}_{\xi,1}\left(\frac{b-\gamma}{c}, 0\right), \frac{b}{c}\right) = \infty.$$

This completes the proof of the theorem. □

Proof of Lemma 3.3 Let $H \in \mathcal{P}_{H_0}(c,\gamma)$. For a given $\boldsymbol{\theta} \in R^p$ let $F = F^{\boldsymbol{\theta}}$ be the distribution of the residual $r(\boldsymbol{\theta}) = y - \boldsymbol{\theta}'\boldsymbol{x}$ under H. Let $S_n(F)$ and $\tilde{S}(F)$ be the M-scale functionals corresponding to ρ_n and $\tilde{\rho}$. Then we have $0 < S_n(F), \tilde{S}(F) < \infty$. According to the definitions (2.7) and (2.8) of $\boldsymbol{T}_n(F)$ and $\tilde{\boldsymbol{T}}(F)$, for the proof of the lemma it is sufficient to show that there exists a subsequence $\{S_{n_k}(F)\}$ of $\{S_n(F)\}$ such that $\lim_{k\to\infty} S_{n_k}(F) = \tilde{S}(F)$. We note that $E_F[\rho_n(u/s)]$ and $E_F[\tilde{\rho}(u/s)]$ are strictly decreasing and decreasing (nonincreasing) in $s > 0$, respectively, and that

$$\lim_{n\to\infty} E_F\left[\rho_n\left(\frac{u}{s}\right)\right] = E_F\left[\tilde{\rho}\left(\frac{u}{s}\right)\right], \quad {}^{\forall}s > 0 \text{ uniformly.} \tag{5.10}$$

Also, by A5 we note that $E_F[\rho_n(u/s)]$ and $E_F[\tilde{\rho}(u/s)]$ are left continuous in s. It is easy to see from the definitions of $S_n(F)$ and $\tilde{S}(F)$ that for $0 < {}^{\forall}\varepsilon < S_n(F) and\ {}^{\forall}n$,

$$E_F\left[\rho_n\left(\frac{u}{S_n(F)-\varepsilon}\right)\right] > E_F\left[\rho_n\left(\frac{u}{S_n(F)}\right)\right] \geq b > E_F\left[\rho_n\left(\frac{u}{S_n(F)+\varepsilon}\right)\right], \tag{5.11}$$

and that for $0 < {}^{\forall}\varepsilon < \tilde{S}(F)$

$$E_F\left[\tilde{\rho}\left(\frac{u}{\tilde{S}(F)-\varepsilon}\right)\right] > E_F\left[\tilde{\rho}\left(\frac{u}{\tilde{S}(F)}\right)\right] \geq b \geq E_F\left[\tilde{\rho}\left(\frac{u}{\tilde{S}(F)+\varepsilon}\right)\right], \tag{5.12}$$

Let $0 < {}^{\forall}\varepsilon < \tilde{S}(F)$. Since

$$E_F\left[\tilde{\rho}\left(\frac{u}{\tilde{S}(F)-\varepsilon}\right)\right] > b,$$

it follows from (5.10) that there exists an integer n_0 such that

$$E_F\left[\rho_n\left(\frac{u}{\tilde{S}(F)-\varepsilon}\right)\right] > b, \quad {}^{\forall}n \geq n_0. \tag{5.13}$$

Hence, noting

$$E_F\rho_n\left(\frac{u}{S_n(F)+\varepsilon}\right) < b, \quad {}^{\forall}n, \tag{5.14}$$

we have

$$\tilde{S}(F) - \varepsilon < S_n(F) + \varepsilon, \quad {}^{\forall}n \geq n_0.$$

This implies that

$$\tilde{S}(F) \leq \liminf_{n\to\infty} S_n(F). \tag{5.15}$$

On the other hand, to show the reverse inequality in (5.15). We note that it follows from the definitions of $S_n(F)$ and $\tilde{S}(F)$ that for any sufficiently small $\varepsilon > 0$,

$$\lim_{n\to\infty} E_F\left[\rho_n\left(\frac{u}{S_n(F)-\varepsilon}\right)\right] = E_F\left[\tilde{\rho}\left(\frac{u}{\tilde{S}(F)-\varepsilon}\right)\right] > b.$$

Since

$$E_F\left[\tilde{\rho}\left(\frac{u}{\tilde{S}(F)+\varepsilon}\right)\right] \le b,$$

for any sufficiently small $\varepsilon > 0$ there exists an integer n_0 such that

$$E_F\left[\rho_n\left(\frac{u}{S_n(F)-\varepsilon}\right)\right] > E_F\left[\rho_n\left(\frac{u}{\tilde{S}(F)+\varepsilon}\right)\right], \qquad {}^{\forall}n \ge n_0.$$

Hence we have

$$S_n(F)-\varepsilon < \tilde{S}(F)+\varepsilon, \quad {}^{\forall}\varepsilon > 0, {}^{\forall}n \ge n_0.$$

This implies that

$$\limsup_{n\to\infty} S_n(F) \le \tilde{S}(F). \tag{5.16}$$

Thus, form (5.15) and (5.16) we obtain

$$\lim_{n\to\infty} S_n(F) = \tilde{S}(F).$$

This completes the proof of the lemma. □

Proof of Theorem 4.1 Let $\xi_k = \{W^k_{s,\theta}\}$ be the element of $\mathcal{F}_{1\lambda}$ such that the density $f^{\xi_k}_{s,\|\theta\|}$ of $\frac{y-\boldsymbol{\theta}'\boldsymbol{x}}{s}$ under $H_0 - W^k_{s,\theta}$ is given by (4.1) with

$$\phi_{\xi_k}(u) \;=\; \begin{cases} \phi(u), & \text{if } \; 0 \le |u| < z_1, \\ \phi(u) - \frac{k}{2(z_2-z_1)}, & \text{if } \; z_1 \le |u| < z_2, \\ \phi(u), & \text{if } \; |u| \ge z_2, \end{cases} \tag{5.17}$$

where k is a constant satisfying $0 < k < \min\left\{\frac{c+\gamma-1}{c}, 2(z_2-z_1)\phi(z_2)\right\}$, $z_1 = z_{\left(\frac{b+k}{2}\right)}$ and $z_2 = z_{\left(\frac{b}{2}-\frac{c+\gamma-1}{2c}\right)}$. It is easy to see that

$$h_{\xi_k}(\tau) \;=\; \begin{cases} 2\left(1-\Phi\left(\frac{1}{\tau}\right)\right), & \text{if } \; 0 \le \tau \le \frac{1}{z_2}, \\ 2(1-\Phi(z_2)) + \frac{k}{z_2-z_1}\left(k-\frac{1}{\tau}\right), & \text{if } \; \frac{1}{z_2} \le \tau \le \frac{1}{z_1}, \\ 2\left(1-\frac{c+\gamma-1}{2c}-\Phi\left(\frac{1}{\tau}\right)\right), & \text{if } \; \tau \ge \frac{1}{z_1}. \end{cases} \tag{5.18}$$

Since $c < 1$, we have

$$h_{\xi_k}\left(\frac{1}{z_2}\right) > \frac{b-\gamma}{c} \quad \text{and} \quad h_{\xi_k}\left(\frac{1}{z_1}\right) < \frac{b}{c},$$

and hence

$$h^{-1}_{\xi_k}\left(\frac{b-\gamma}{c}\right) \;=\; \frac{1}{\Phi^{-1}\left(1-\frac{b-\gamma}{2c}\right)} = h^{-1}_{\xi^*}\left(\frac{b-\gamma}{c}\right), \tag{5.19}$$

$$h^{-1}_{\xi_k}\left(\frac{b}{c}\right) \;=\; \frac{1}{\Phi^{-1}\left(1-\frac{b+\gamma+c-1}{2c}\right)} = h^{-1}_{\hat{\xi}}\left(\frac{b}{c}\right). \tag{5.20}$$

Therefore we obtain $\underline{B}_{\boldsymbol{T}_b}(c,\gamma) = \overline{B}_{\boldsymbol{T}_b}(c,\gamma)$ from Corollary 4.1 and the fact $\xi_k \in \mathcal{F}_{1\lambda}$. This completes the proof of the theorem. □

Acknowledgments. This research was supported by the Hori Information Science Scholarship Foundation and by Nanzan Pache Research Subsidy I-A, 2001. The authors would also like to thank Dr.Itsuro Kakiuchi at Kobe University and Dr.Shinichi Matsuda at Nanzan University for their helpful comments.

References

Ando,M. and Kimura,M.(2001). A characterization of the neighborhoods defined by certain special capacities and their applications to bias-robustness of estimates. submitted.

Berrendero,J.R. and Zamar,R.H.(2001). Maximum bias curves for robust regression with non-elliptical regressors. *Ann. Statist.*, **29**, 224-250.

Croux,C., Rousseeuw,P.J. and Bael,A.V.(1996). Positive-breakdown regression by minimizing nested scale estimators. *J. Statist. Plann. Inference.*, **53**, 197-235.

Davies, L.(1990).The asymptotics of S-estimators in the linear regression model. *Ann. Statist.*, **18**, 1651-1675.

Davies, L.(1993). Aspects of robust linear regression. *Ann. Statist.*, **21**, 1843-1899.

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. and Simpson,D.G.(1993). Lower bounds for contamination bias: Globally minimax versus locally linear estimation. *Ann. Statist.*, **21**, 314-337.

Hennig, C.(1995). Efficient high-breakdown-point estimators in robust regression: which function to choose? Statistics & Decisions **13**,221-241.

Hössjer, O.(1992). On the optimality of S-estimators. *Statist. Probab. Lett.*, **14**, 413-419.

Hössjer,O., Croux,C. and Rousseeuw,P.J.(1994). Asymptotics of Generaized S-Estimator. *J. Multivariate Anal.*, **51**, 148-177.

Huber, P.J.(1973). Robust regression: Asymptotics, Conjecture and Monte Carlo. *Ann. Statist.*, **1**, 799-821.

Martin, R.D., Yohai, V.J. and Zamar, R.H.(1989). Min-max bias robust regression. *Ann. Statist.*, **17**, 1608-1630.

Rieder, H.(1977). Least favorable pairs for special capacities. *Ann. Statist.*, **5**, 909-921.

Rousseeuw, P.J.(1984). Least median of squares regression. *J. Amer. Statist. Assoc.*, **79**, 871-880.

Rousseeuw, P.J. and Yohai, V.(1984).Robust regression by means of S-estimators. *Robust and Nonlinear Time Series Analysis.* Lecture Notes in Statistics, **26**, 256-272, Springer, New York.

Rousseeuw,P.J.(1985). Multivariate estimation with high breakdown point. *Math. Stat. and Appl.*, Vo 1.B, 282-297.

Yohai,V.J.(1987). High breakdown point and high efficiency robust estimates for regression. *Ann. Statis.* **15**. 642-656.

Yohai, V.J. and Zamar, R.H.(1988). High breakdown-point estimates of regression by means of the minimization of an efficient scale. *J. Amer. Statist. Assoc.*, **83**, 406-413.

Yohai,V.J. and Zamar,R.H.(1993). A minimax-bias property of the least α-quantile estimates. *Ann. Statist.*, **21**, 1824-1842.

Masakazu Ando
Graduate School of Business Administration, Nanzan University.
18 Yamazato-cho, Showa-ku, Nagoya 466-8673, JAPAN
E-mail: anne@nanzan-u.ac.jp

Miyoshi Kimura
Department of Mathematical Sciences, Nanzan University.
27 Seirei-cho, Seto-shi, Aichi, 489-0863, JAPAN
E-mail: kimura@ms.nanzan-u.ac.jp